

MÔ HÌNH HÓA CÁC QUÁ TRÌNH XỬ LÝ NƯỚC THẢI BẰNG MẠNG NƠON NHÂN TẠO

Nguyễn Kỳ Phùng, Nguyễn Khoa Việt Trường
Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

1. MỞ ĐẦU

Trong những năm gần đây mạng nơon nhân tạo ANN (Artificial Neural Network) ngày càng được ứng dụng rộng rãi trong công tác dự báo, mô phỏng ở rất nhiều lĩnh vực: tài chính, năng lượng, y học, tài nguyên nước và khoa học môi trường. Đặc biệt là trong lĩnh vực kỹ thuật môi trường ANN ngày càng chứng tỏ được vai trò trong mô phỏng các quá trình xử lý phức tạp mà các công cụ mô hình hóa thông thường hay bộc lộ những nhược điểm của nó.

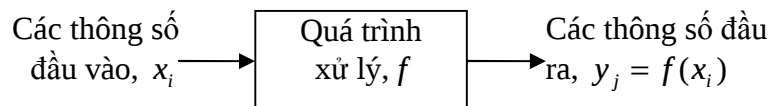
Ở Việt Nam hiện nay, mặc dù đã có những hệ thống xử lý nước thải được điều khiển tự động hóa nhưng quá trình vận hành các hệ thống này đòi hỏi kinh nghiệm khá cao của các nhân viên vận hành. Công việc vận hành đòi hỏi phải tiến hành các thí nghiệm thường xuyên rất mất nhiều thời gian và tốn kém. Nếu mạng nơon được ứng dụng vào công tác vận hành các hệ thống thì sẽ hứa hẹn một hiệu quả cao hơn và đáng tin cậy hơn...

Bài báo này trước hết tổng quan được vấn đề mô hình hóa các quá trình xử lý nước thải, tiếp đó là giới thiệu sơ bộ lý thuyết của mạng nơon nhân tạo. Phần trọng tâm của bài báo tập trung vào ứng dụng cho một nhà máy xử lý nước thải cụ thể. Cuối cùng là kết quả và thảo luận.

2. MÔ HÌNH HÓA CÁC QUÁ TRÌNH XỬ LÝ NƯỚC THẢI

Muốn mô phỏng một hệ thống kỹ thuật, người ta phải tìm cách nào để mô tả được quy luật hoạt động của hệ thống đó. Hay nói cách khác, người ta phải cố gắng tìm được mối liên hệ giữa các thông số đầu vào và đầu ra của hệ thống.

Như ta đã biết, hệ thống xử lý nước thải cũng như bất kỳ hệ thống kỹ thuật nào khác đều bao gồm nhiều công trình đơn vị trong đó. Mỗi công trình đều có một chức năng riêng, tất cả được kết nối thành một hệ thống và cùng nhau thực hiện một chức năng tổng quát, đối với hệ thống xử lý nước thải là: biến đổi nước thải thành nước sạch theo một tiêu chuẩn nào đó. Ta có thể sơ đồ hóa các công trình đơn vị của hệ thống xử lý nước thải như sau:



Hình 1. Sơ đồ đại diện cho công trình đơn vị trong hệ thống xử lý nước thải

Trong đó:

x_i đại diện cho các thông số đầu vào như lưu lượng Q, BOD, pH, NH_4^+ ...

Tương tự y_j đại diện cho các thông số đầu ra như BOD, COD, pH...

f ở đây đại diện các quá trình xử lý, có thể là bể lắng cát, bể lắng sơ bộ, bể aerotank, bể lắng cấp hai, bể khử trùng...

Theo cách diễn đạt như trên ta có thể xem mỗi công trình đơn vị là một hàm số nào đó chứa đựng mối liên hệ giữa các thông số đầu vào và đầu ra. Và ta cũng có thể xem cả hệ thống là một hàm số tổng hợp của những hàm số con này. Trong lĩnh vực xử lý nước thải, hiện nay người ta cũng cố gắng xây dựng một số lý thuyết để tính toán nhưng hầu hết còn ở mức độ rất đơn giản, điều kiện tính toán thường lý tưởng và kết quả thu được chỉ mang tính chất gần đúng, ước lượng.

Lý do là bản chất vấn đề rất phức tạp, hiệu quả của mỗi công trình xử lý phải phụ thuộc vào rất nhiều yếu tố mà ta không thể xét hết được. Thêm vào đó chất lượng nước thải, điều kiện môi trường chứa đựng những thông số rất khó kiểm soát.

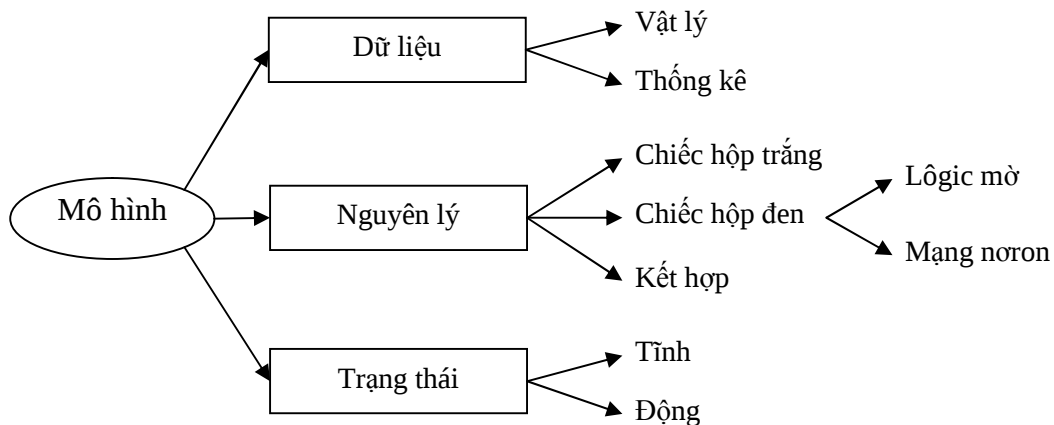
Nhìn một cách tổng quát các công cụ mô hình hóa được sử dụng để mô phỏng các quá trình xử lý nước thải, ta có thể phân loại như sau:

- Theo đặc điểm của dữ liệu được sử dụng: ta có mô hình xác định (Deterministic), còn gọi là mô hình vật lý và mô hình thống kê (Stochastic).
- Xét về mặt phương pháp luận, người ta chia làm mô hình theo nguyên lý chiếc hộp trắng “White box” và nguyên lý chiếc hộp đen “Black box”. Ngoài ra còn có dạng mô hình kết hợp hai nguyên lý trên gọi là mô hình Hybrid.

Nguyên lý chiếc hộp trắng “White box”: trong mô hình theo nguyên lý này, người ta cố gắng mô tả tất cả các quá trình xảy ra bên trong hệ thống bằng các phương trình toán học. Chẳng hạn như trong một hệ thống xử lý nước thải bằng phương pháp sinh học, các mô hình dựa trên nguyên lý này sẽ mô phỏng nhiều nhất có thể các quá trình thủy lực, sinh hóa ... bằng các phương trình toán.

Nguyên lý chiếc hộp đen “Black box”: ngược lại với nguyên lý trên, nguyên lý “Black box” không quan tâm đến những gì xảy ra bên trong hệ thống, coi nó như là một “chiếc hộp đen”. Các mô hình dựa trên nguyên lý này chỉ quan tâm đến giá trị của các thông số ở đầu vào và đầu ra của hệ thống. Và mối liên hệ giữa các thông số này được thiết lập dựa trên các công cụ thống kê.

- Xét theo trạng thái, người ta chia mô hình ra hai loại: tĩnh (static) và động (dynamic).



Hình 0. Sơ đồ phân loại các công cụ mô hình hóa.

Dựa trên cơ sở nào để lựa chọn loại mô hình cần thiết để giải quyết vấn đề đặt ra? Tác giả Baba cùng cộng sự đã xây dựng một sơ đồ cây mô tả các nguyên tắc của quá trình lựa chọn như hình 2.3 (Zvi Boger, 1992) [**Error! Reference source not found.**].

Qua sơ đồ ta có thể phân tích được một số điểm sau:

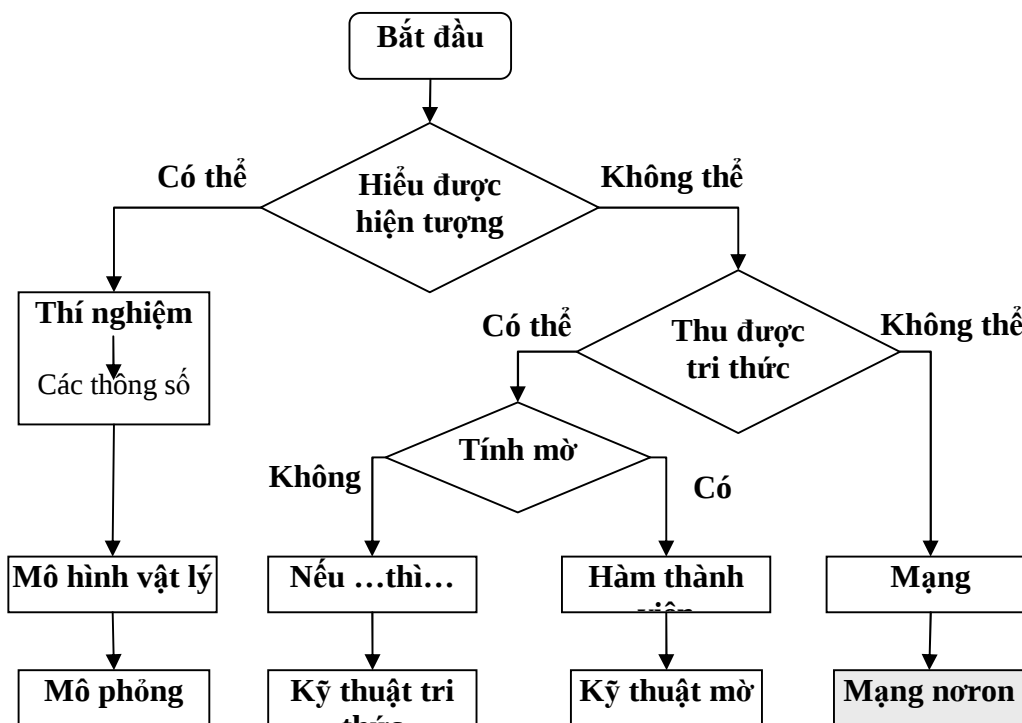
- Đối với một quá trình mà ta biết rõ được hiện tượng, để mô phỏng nó người ta tiến hành các thí nghiệm với các thông số đã nhận thức được, từ kết quả thí nghiệm này ta có thể xây dựng các mô hình vật lý (physical model), sau đó sử dụng mô hình này để mô phỏng quá trình đó.
- Đối với một quá trình mà ta chưa hiểu được hiện tượng, hoặc hiện tượng rất phức tạp đến nỗi ta không mô tả được, ta có hai hướng để quyết định. Nếu ta có thể thu được “tri thức”

(knowledge) từ quá trình đó, ta có thể lựa chọn công cụ mờ (Fuzziness) để giải quyết. Còn trong trường hợp ta không thể thu được “tri thức” từ bản chất vấn đề, mạng (network) (ở đây là mạng nơron) sẽ là một giải pháp tốt hơn nhiều.

Qua quá trình phân tích từ hai sơ đồ trên, ta có thể thấy được mạng nơron là loại mô hình mà theo đặc điểm của dữ liệu nó thuộc về mô hình thống kê, xét về mặt phương pháp luận nó là mô hình theo nguyên lý chiếc hộp đen, nguyên lý mà ta không quan tâm đến các quá trình xảy ra bên trong hệ thống mà chỉ quan tâm đến dữ liệu đầu vào và đầu ra của hệ thống đó. Theo trạng thái, tùy theo ứng dụng và cấu trúc của từng mạng nơron cụ thể mà ta có thể xếp nó vào loại tĩnh hoặc động.

Và ta cũng có thể nhận thấy được rằng mạng nơron là một giải pháp rất thích hợp cho mô phỏng các quá trình xử lý nước thải vốn được xem là bản chất rất phức tạp, khó kiểm soát. Thêm vào đó sự kết hợp của công cụ mạng nơron với các công cụ mô hình khác chẳng hạn các mô hình vật lý, logic mờ... như hứa hẹn cho ra đời những công cụ mô phỏng mạnh, hiệu quả trong tương lai.

Mạng nơron là công cụ mô hình và điều khiển được sử dụng rộng rãi trong rất nhiều quá trình công nghiệp phi tuyến, chúng có thể là các mô hình on-line (Ngia and Sj .oberg, 2000) [[10]] hoặc off-line (Lightbody and Irwin, 1997 [[9]]; Bloch et al., 1997) [[6]]. Các hệ thống nhúng mờ cũng được sử dụng rộng rãi, có thể kể đến các mô hình on-line như (Fink et al., 2001 [[6]]) và một số mô hình off-line như (Kovacevic and Zhang , 1997 [[8]]; Zhang and Kovacevic, 1998 [[12]]; Vieira and Mota, 2003 [[11]]).



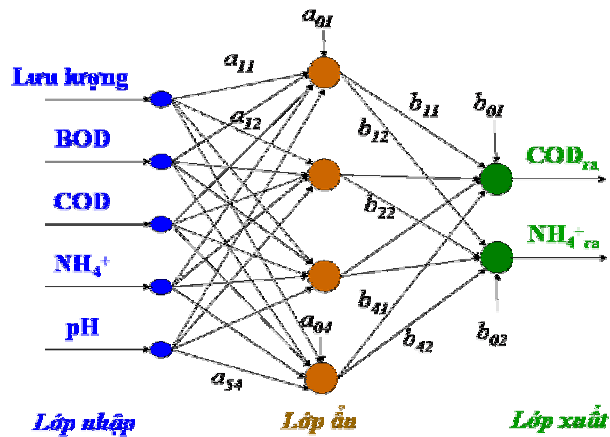
Hình 3. Lựa chọn các kỹ thuật mô hình hóa cho các hệ thống kỹ thuật

3. GIỚI THIỆU MẠNG NƠN NHÂN TẠO

Mạng nơron nhân tạo sử dụng nguyên lý tính toán song song bao gồm nhiều quá trình tính toán đơn giản được kết nối với nhau. Trong mỗi quá trình này, các phép tính được thực hiện rất đơn giản, do một nơron đảm trách. Nhưng chính những nơron đơn giản này lại có thể giải quyết

được những nhiệm vụ rất phức tạp khi chúng được kết nối, tổ chức với nhau theo một cách hợp lý nào đó.

Thực ra, nền tảng của các mạng ANN được đưa ra vào những năm của thập kỷ 50 nhưng mãi đến đầu thập kỷ 90, chúng mới thật sự được chấp nhận rộng rãi và trở thành công cụ hữu ích. Lý do chính là con người đã vượt qua được một số rào cản về lý thuyết cũng như sự phát triển mạnh mẽ về khả năng của phần cứng máy tính. Thuật ngữ “nhân tạo (Artificial)” thực ra được dùng để chỉ công cụ tính toán bằng mạng nơron là sản phẩm trí tuệ của con người chứ không phải mạng nơron sinh học ở bộ não người. Một điều hiển nhiên rằng quá trình tìm hiểu bộ não người có tính chất quyết định quá trình phát triển của các mạng ANN. Tuy vậy, khi so sánh với bộ não người, cơ chế hoạt động của mạng ANN hiện nay còn ở mức độ rất đơn giản. Thêm vào đó mạng ANN thường được đề cập như một mạng kết nối khi khả năng tính toán được nhấn mạnh hơn là tính chính xác về mặt sinh học. Nói cách khác, tính kết nối giúp mạng nơron thực hiện nhiệm vụ của mình chứ không phải cố gắng mô phỏng chính xác phần nào đó của một quá trình sinh học [[3]].

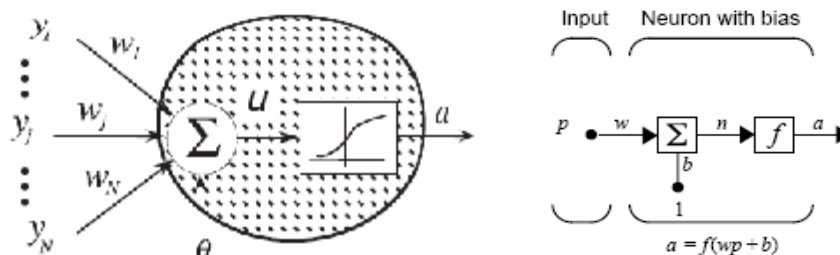


Hình 4. Sơ đồ minh họa một mạng nơron [[1]]

3.1. Nguyên tắc hoạt động của mạng nơron

Mạng nơron bao gồm nhiều đơn vị được liên kết với nhau theo một cách nào đó và cho phép chúng trao đổi thông tin với nhau. Các đơn vị này được xem là một nơron, hay một nút, là những bộ phận xử lý rất đơn giản. Khả năng tính toán của từng đơn vị này rất hạn chế bao gồm một phép cộng các tín hiệu đầu vào và một hàm truyền tính toán tín hiệu đầu ra từ giá trị của các tín hiệu đầu vào đã được cộng gộp đó.

Các tín hiệu đầu ra (output signals) này có thể được gửi đến các đơn vị (nơron) khác bằng một hệ số nào đó gọi là trọng số. Các trọng số này làm tăng cường hay giảm thiểu tín hiệu mà các nơron trao đổi cho nhau. Sơ đồ của một nơron đơn vị có thể được minh họa bằng hình vẽ sau [[5]]:



Hình 5. Mô hình một nơron đơn vị (a) McCulloch và Pitt, (b) biểu diễn toán học

Mặc dù nguyên tắc tính toán của một nơron đơn vị rất đơn giản, nhưng khi được kết hợp với nhau theo một cấu trúc nào đó, nó có khả năng xấp xỉ những hàm phi tuyến rất phức tạp.

3.2. Phát triển mô hình mạng nơron

Khi phát triển một mô hình mạng nơron người ta thường thực hiện theo các bước sau đây:

**Hình 6.** Các bước phát triển mô hình

3.3. Ứng dụng mô hình mạng nơron nhân tạo

Sau khi một mạng đã được xây dựng và kiểm chứng, ta có thể đưa ứng dụng theo các hướng sau:

- Sử dụng mô hình cho công tác nghiên cứu

Một mạng nơron đã được luyện với một cơ sở dữ liệu đầy đủ sẽ là một công cụ rất tốt cho công tác nghiên cứu. Bản thân mạng đã chứa đựng nhiều thông tin về quá trình mà nó được "học" nên ta có thể sử dụng nó khảo sát mối liên hệ giữa các thông số trong quá trình xử lý nước thải vốn được xem là phức tạp, điều này cũng hứa hẹn góp phần hoàn thiện hơn các lý thuyết của các quá trình xử lý nước thải hiện nay.

- Sử dụng mô hình cho thiết kế

Trong quá trình thiết kế bất cứ hệ thống nào, các quá trình xử lý thường được ước tính thông qua các công cụ mô phỏng. Các mạng nơron sau khi đã được luyện tỏ ra rất hữu ích khi thiết kế các hệ thống tự động mà nó đã được luyện. Mạng sẽ cung cấp cho người thiết kế các thông tin hữu ích hỗ trợ quá trình ra quyết định và chọn lựa các thông số vận hành của phù hợp với các thông số đầu vào công cụ hệ thống cần phải thiết kế.

- Sử dụng mô hình cho tối ưu hóa hệ thống

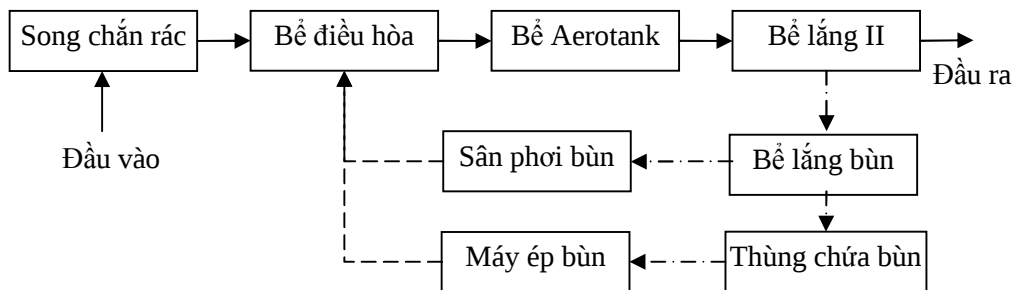
Tối ưu hóa các quá trình có thể hiểu ở các góc độ khác nhau. Tối ưu hóa quá trình ngoại tuyến (off-line) nghĩa là ta mô phỏng quá trình sao cho chúng được vận hành tối ưu, kết quả sẽ được sử dụng và kiểm chứng trên các hệ thống full-scale. Tối ưu hóa quá trình trực tuyến (on-line) được hiểu nôm na là quá trình tối ưu hóa thực hiện trên các hệ thống kiểm soát chất lượng của các hệ thống xử lý đang hoạt động. Cho dù là ngoại tuyến hay trực tuyến thì mạng đã được luyện đóng một vai trò rất quan trọng. Ta có thể kết hợp mạng nơron của ta vào các hệ thống kiểm soát khác như "Model Predict Control (MPC)", "Model Based control (MBC)" hoặc bất cứ kỹ thuật kiểm soát nào khác.

4. ÁP DỤNG CỤ THỂ

Việc đầu tiên là xây dựng mô hình để chạy các số liệu thực tế. Nghiên cứu đã xây dựng được chương trình trên ngôn ngữ MatLab với các chức năng: xử lý sơ bộ số liệu, tạo mạng nơron, luyện mạng và kiểm chứng mạng, tối ưu hóa mạng để xác định số nút ẩn và số vòng lặp tối ưu, và sau cùng là hiển thị kết quả bằng đồ thị.

Để minh họa ứng dụng mạng nơron nhân tạo, kiểm tra hiệu quả của mô hình đã xây dựng được, nghiên cứu đã sử dụng số liệu của một số trạm xử lý nước thải trong và ngoài nước. Bài báo này chỉ trình bày kết quả của trường hợp nhà máy sữa Cô Gái Hà Lan. Kết quả của các trạm xử lý khác bạn đọc có thể tham khảo ở tài liệu [[1]].

4.1. Hệ thống xử lý nước thải nhà máy sữa Cô Gái Hà Lan



Hình 7. Sơ đồ hệ thống xử lý nước thải nhà máy sữa Cô Gái Hà Lan

4.2. Thu thập và tổ chức số liệu

Số liệu được thu thập ở các vị trí đầu vào và đầu ra của hệ thống với những đặc trưng như sau (sẽ được trình bày ở phần phụ lục):

- Các thông số đầu vào bao gồm pH, lưu lượng Q, nhu cầu ôxy hóa học COD và chất rắn lơ lửng SS.
- Thông số đầu ra: COD
- Số mẫu quan sát: 88

Dữ liệu sau khi thu thập được tổ chức dưới dạng 1 x 3 cell như sau:

CGHLip = {4x1 cell} [4x67 double] [4x21 double]

CGHLoPCOD = {1x1 cell} [1x67 double] [1x21 double]

Dữ liệu đầu vào: gồm có 4 thông số pH, lưu lượng Q, COD và SS. Dữ liệu được chia làm hai tập: tập để luyện mạng là một ma trận [4x67] và tập để kiểm chứng là một ma trận [4x21].

Dữ liệu đầu ra: có một thông số là COD. Dữ liệu được chia làm hai tập: tập để luyện mạng là một ma trận [1x67] và tập để kiểm chứng là một ma trận [1x21].

4.3. Kết quả

Tiêu chí được chọn trong quá trình luyện mạng và kiểm chứng mạng là MAE và MRSE.

Sai số MAE (Mean Absoluted Errors): sai số tuyệt đối trung bình. Được tính toán như sau:

$$MEA = \frac{\sum_{i=1}^n \frac{e_i}{T_i}}{n}, \text{ trong đó:}$$

- e_i là sai số dự báo của quan sát thứ i
- T_i là giá trị thực tế của quan sát thứ i

- n là số quan sát.

Sai số RMSE (Root Mean Squared Errors): sai số bình phương trung bình gốc. Được tính toán như sau:

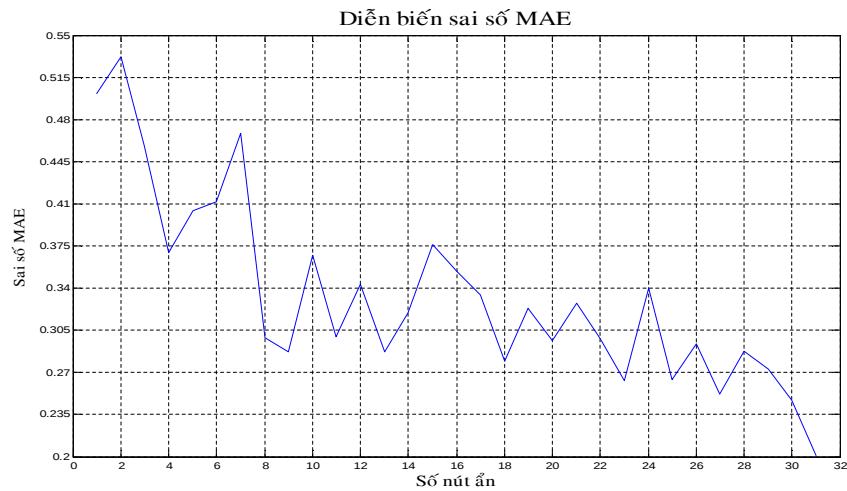
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}}, \text{ trong đó:}$$

- e_i và n có ý nghĩa giống hệt như trường hợp MAE.

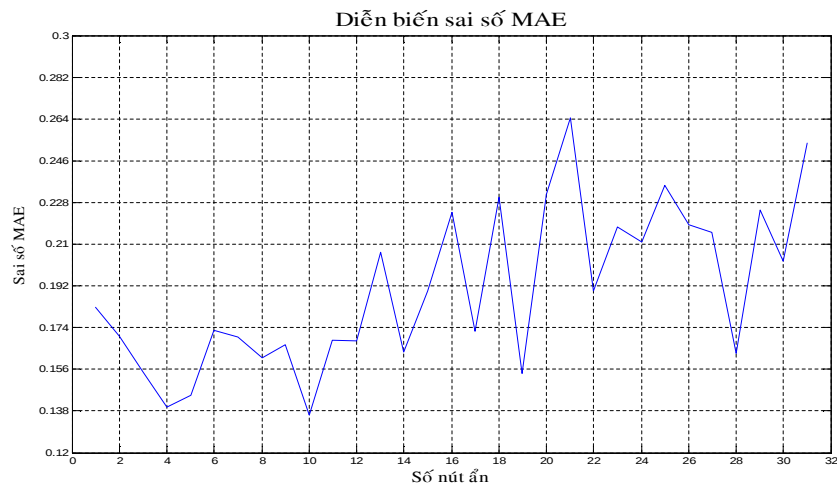
Trong cả hai trường hợp, $e_i = P_i - T_i$ trong đó P_i là giá trị dự báo của quan sát thứ i .

Trước hết ta tiến hành tối ưu hóa mạng và chọn được số bước lặp tối ưu là 13. với tiêu chí MAE, kết quả luyện mạng và kiểm chứng mạng với 13 vòng lặp như sau:

- Trong quá trình luyện mạng mối liên hệ giữa MAE và số nút ẩn như sau:

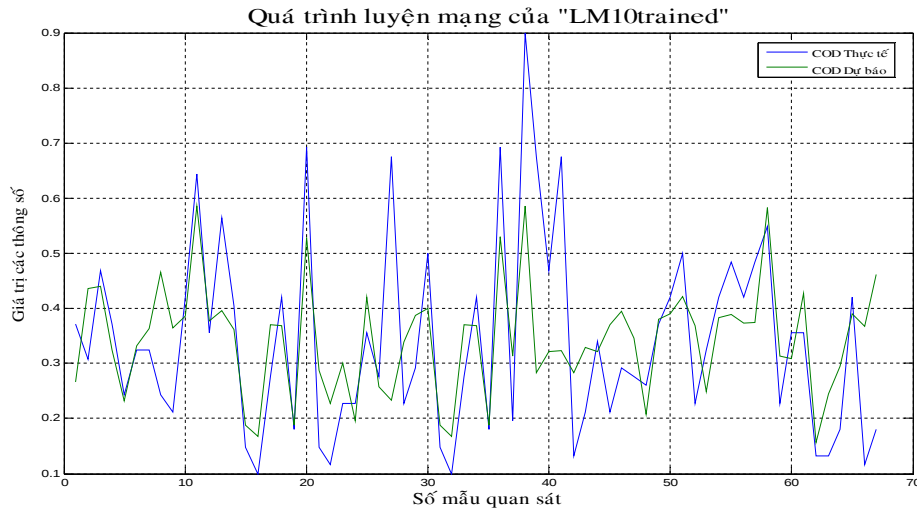


- Khảo sát quá trình kiểm chứng mạng



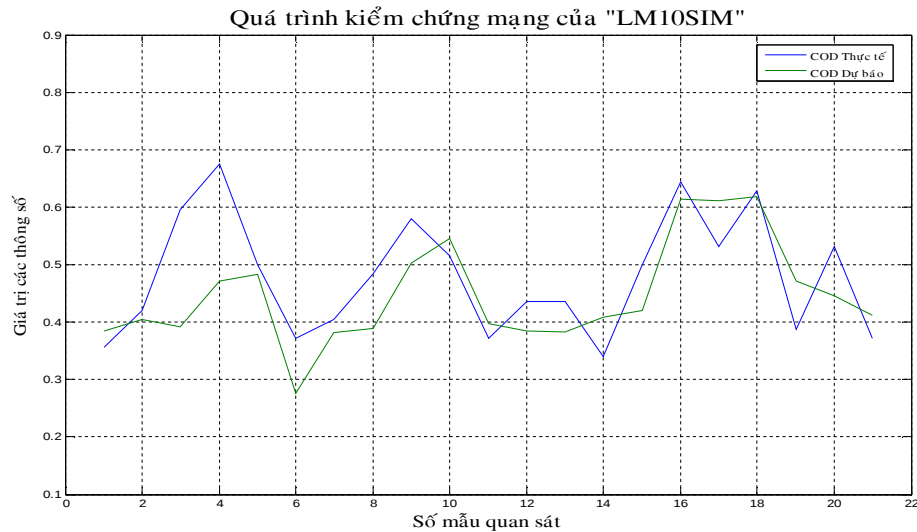
Qua số liệu từ hai đồ thị trên ta có thể chọn được cấu trúc mạng tốt nhất là mạng có 10 nút ẩn cho sai số nhỏ nhất. Sau đây là minh họa cụ thể.

- **Quá trình luyện mạng:**



Trong trường hợp này cho sai số $MAE = 0.367391$ và $RMSE = 0.135688$.

Quá trình kiểm chứng mạng:



Trong trường hợp này cho sai số $MAE = 0.136055$ và $RMSE = 0.084701$.

4.4. Thảo luận

Qua quá trình luyện mạng nơron cho tập dữ liệu thu được từ các nhà máy xử lý nước thải công ty sữa Cô Gái Hà Lan, ta có thể rút ra được một số kết quả sau:

Trong suốt quá trình luyện mạng, sai số có xu hướng giảm khi số nút ẩn tăng nhưng trong quá trình kiểm chứng mạng thì sai số sẽ giảm đến một giá trị nút ẩn tối ưu nào đó, sau đó sẽ tăng trở lại, qua đó ta có thể thấy được số nút ẩn tối ưu của mạng là 10.

Sử dụng kỹ thuật luyện mạng có so sánh một cách tự động sẽ tránh được tình trạng quá khớp của mạng, giảm được đáng kể thời gian thử và sai để xác định ở thể hệ luyện nào mạng sẽ quá khớp.

Nhờ áp dụng thuật toán tối ưu hóa luyện mạng liên tục lặp nhiều lần mà kết quả thu được ổn định và tốt hơn nhiều. Đặc điểm của thuật toán mạng nơron là có thể cho ra những kết quả khác nhau trong những lần luyện khác nhau, do đó có thể cho ra sai số không ổn định. Với tiến trình lặp lại nhiều lần, mỗi lần sẽ chọn mạng cho sai số nhỏ nhất. Nhờ vậy mà khoảng biến thiên sai số sẽ ổn định mặc dù số nút ẩn của mạng cho kết quả tốt nhất có thể thay đổi.

5.KẾT LUẬN

Kết quả về mặt cơ bản đã tiến hành nghiên cứu ứng dụng kỹ thuật mạng nơron để dự báo chất lượng của một hệ thống xử lý nước thải cả về lý thuyết lẫn áp dụng thực tiễn.

Kết quả đạt được rất khả quan, mạng nơron có thể dự báo kết quả đầu ra của hệ thống xử lý nước thải với mức độ chính xác hoàn toàn có thể chấp nhận được. Sai số MAE = 0.136055 và RMSE = 0.084701.

Đã xây dựng chương trình ứng dụng mạng nơron cho dự báo chất lượng đầu ra của một hệ thống xử lý nước thải. Cùng với thuật toán tối ưu hóa mạng nơron khi cho số nút ẩn thay đổi để tìm ra cấu trúc mạng tốt nhất, chương trình đã thể hiện rõ ưu so với chương trình mạng nơron của MatLab.

Thuật toán tối ưu hóa quá trình luyện mạng là một bước cải tiến so với các chương trình ứng dụng mạng nơron thông thường, chẳng hạn như MatLab. Với quá trình lặp lại nhiều lần và ghi nhận những mạng cho kết quả tốt nhất sau mỗi lần lặp, ta có thể chọn được mạng cho kết quả tốt hơn và sai số ổn định hơn.

Trong các nghiên cứu tiếp, nhóm tác giả sẽ tiếp cận và đưa thuật toán di truyền vào để nâng cao độ chính xác của bài toán.

TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Khoa Việt Trường, *Ứng dụng mạng nơron nhân tạo để dự báo đầu ra của hệ thống xử lý nước thải*, Luận Văn Tốt Nghiệp Thạc Sĩ, tháng 03 năm 2006.
- [2]. Nguyễn Kỳ Phùng, Đậu Thị Dung. *Ứng dụng mạng nơron tính toán và dự đoán đầu ra của hệ thống xử lý nước thải nhà máy sữa Cô gái Hà Lan*. Tạp chí Khí tượng thủy văn, số 551,11/2006.
- [3]. Callan, Robert, *Essence of neural networks*, Prentice Hall Europe, (1999).
- [4]. The Mathworks, *Matlab programming*, version 7, (2004).
- [5]. The Mathworks, *Neural Network Toolbox User's Guide for use with Matlab*, version 4, (2004).
- [6]. Bloch, G., Sirou, F., Eustache, V., Fatrez, P., *Neural intelligent control for a Stell Plant*. IEEE Transactions on Neural Networks 8 (4), 910–918., (1997).
- [7]. Fink, A., Nelles, O., Fischer, M., Isermann, R., *Nonlinear adaptive control of a heat exchanger*, International Journal of Adaptive Control and Sig.1 Proc. 15 (8), 883–906., (2001).
- [8]. Kovacevic, R., Zhang, Y.M., *Neurofuzzy model-based weld fusion state estimation*, IEEE Control Systems Magazine 17 (2), 30–42., (1997).
- [9]. Lightbody, G., Irwin, G.W., *Nonlinear control structures based on embedded neural system models*, IEEE Transactions on Neural Networks 8 (3), 553–567, (1997).

- [10]. Ngia, L., Sjoberg, J., *Efficient training of neural nets for nonlinear adaptive filtering using a recursive Levenberg–Marquardt algorithm*, IEEE Transactions in Signal Processing 48 (7), 1915–1927., (2000).
- [11]. Vieira, J., Mota, A., *Smith predictor based neural fuzzy controller applied in a water gas heater that presents a large time-delay and load disturbances*. Proceedings of IEEE International Conference on Control Application, Vol. 1, pp. 362–367., (2003).
- [12]. Zhang, Y.M., Kovacevic, R., *Neurofuzzy model based control of weld fusion zone geometry*. IEEE Transactions on Fuzzy Systems 6 (3), 389–401., (1998).