

ƯỚC LƯỢNG THÔNG SỐ TÀI NGUYÊN VÀ MÔI TRƯỜNG THEO THỜI GIAN ĐỐI VỚI CÁC ĐỊA ĐIỂM CÓ DỮ LIỆU TƯƠNG QUAN BẰNG CÁCH ỨNG DỤNG MẠNG NƠ RON ĐỂ XỬ LÝ TÍN HIỆU SỐ

Trần Cảnh Dương

Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Trong quá trình tìm giải pháp cho các vấn đề kỹ thuật có lúc ta sẽ cần tìm giá trị cực tiểu hay cực đại cho các hàm mục tiêu nhiều chiều. Thông thường việc xác định giá trị tối ưu sẽ dẫn đến việc tìm nghiệm của hệ phương trình phi tuyến. Tài nguyên và môi trường có nhiều thông số, tuy vậy việc quan trắc, thống kê, lưu trữ dữ liệu cho các cơ quan, địa phương không đầy đủ cho nhiều nơi trong khoảng thời gian dài. Một số công trình cần thống kê dữ liệu đủ vị trí và thời gian mới có thể thực hiện do đó cần đo đạc mất nhiều thời gian, công sức và kinh phí. Bài báo này sử dụng phương pháp tìm giá trị cực trị đa biến và đề xuất phương pháp ứng dụng mạng nơ ron để xử lý tín hiệu số từ đó ước lượng biểu đồ các thông số tài nguyên và môi trường theo thời gian đối với các địa điểm có dữ liệu tương quan. Kết quả của bài báo cho thấy ta có thể sử dụng giải pháp nêu trên để rút ngắn thời gian đo đạc các tham số tài nguyên và môi trường dựa trên các số liệu đo đạc quan trắc tại một điểm lân cận với vị trí đã chọn hoặc một vị trí khác có số liệu tương quan. Như vậy, thời gian và kinh phí đo đạc dữ liệu được giảm, trong khi đó sự vận hành của các công trình vẫn được đảm bảo với một sai số cho phép.

Từ khóa: Mạng nơ ron; Xử lý tín hiệu số; Thông số tài nguyên và môi trường; Giá trị cực trị đa biến; Tham số học thích nghi.

Abstract

Estimating resources and temporal environment parameters for locations with correlated data using neural networks to process digital signals

In the process of finding solutions to technical problems, sometimes, it is needed to find minimum or maximum values for multi-dimensional target functions. Usually determining the optimal value will lead to finding solutions of the system of nonlinear equations. There are many natural resources and the environment parameters, however, the monitoring, and storage of these parameters for for different agencies are not organized and effective spatially and temporally. A number of works need sufficient statistical data for various location and time period. Therefore, more funds and effort are needed for extra measurement to collect data. This research uses the method of finding multivariate extreme values and proposes the method of applying neural networks to process digital signals in order to estimate the graph of natural resource and environmental parameters over time for the locations that have correlated data. The results show that the above-mentioned solution can be used to shorten the measurement time of natural resources and environment parameters based on the measurement data of the point adjacent to the selected location or of another location with correlated data. Thus, time and finance for data measurement are reduced while the works can still be conducted with an acceptable errors.

Keywords: Neural networks; Digital signals processing; Natural resources and environment parameters; Multivariate extreme values; Adaptive learning parameters.

1. Giới thiệu bài toán tìm giá trị cực trị đa biến

Khi giải bài toán tìm giá trị tối ưu ta quan tâm điểm cực trị của hàm mục tiêu [3]. Bài báo này đưa ra ví dụ tìm giá trị cực tiểu của một hàm bằng đồ thị hoặc các chòm lệnh. Để biểu diễn đồ thị của hàm

$$Z = X^5 + Y^4 - X^2 - 2Y^2 + \sin(XY)$$

$Z = X^5 + Y^4 - X^2 - 2Y^2 + \sin(XY)$ ta thực hiện lệnh như sau:

```
>> x = -2:0.05:2;
```

```
y = -2:0.05:2;
```

```
[X,Y] = meshgrid(x,y);
```

```
Z = X.^5 + Y.^4 - X.^2 - 2*Y.^2 +
```

```
sin(X.*Y.*1);
```

```
>> figure
```

```
>> surf(X,Y,Z);
```

```
>> axis([-2,2,-2,2,-5,12])
```

```
>> title('Ham muc tieu Z')
```

```
>> xlabel('x1')
```

```
>> ylabel('x2')
```

```
>> zlabel('f(x1,x2)')
```

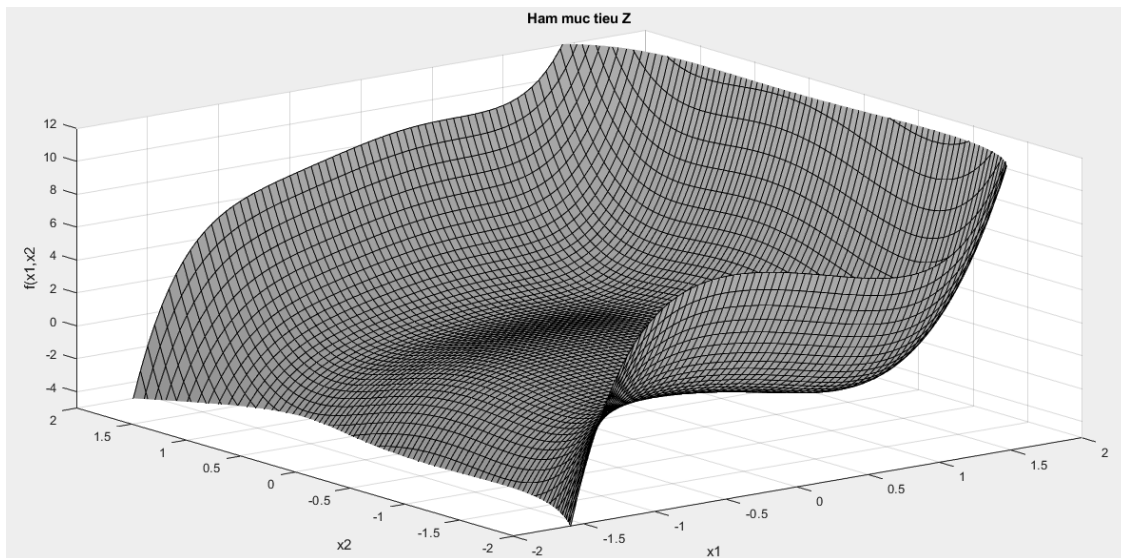
```
View (-40, 35)
```

Đồ thị của hàm

$$Z = X^5 + Y^4 - X^2 - 2Y^2 + \sin(XY)$$

$$Z = X^5 + Y^4 - X^2 - 2Y^2 + \sin(XY)$$

được biểu thị ở Hình 1.



Hình 1: Đồ thị của hàm $Z = X^5 + Y^4 - X^2 - 2Y^2 + \sin(XY)$

Ta tạo script đối với hàm như sau:

```
function F = function9(x)
```

```
F = 2*x(1)^5 + x(2)^4 - x(1)^2 - 2*x(2)^2 + sin((1)*x(2));
```

```
end
```

Sau đó ta thực hiện ở cửa sổ lệnh:

```
>> options = optimset('fminunc');
```

```
>> options = optimset(options, 'Display', 'iter', 'LargeScale', 'off');
```

```
>> x0 = [1 -1];
```

```
>> [x, fval] = fminunc('function9', x0, options)
```

Ta có kết quả sau:

Nghiên cứu

		First-order		
Iteration	Func-count	f(x)	Step-size	Optimality
0	3	-0.841471		8
1	6	-1.85652	0.125	0.114
2	9	-1.85708	1	0.0333
3	12	-1.85714	1	0.000463
4	15	-1.85714	1	0.000482
5	42	-6.05271e+20	4.84276e+07	2.43e+17

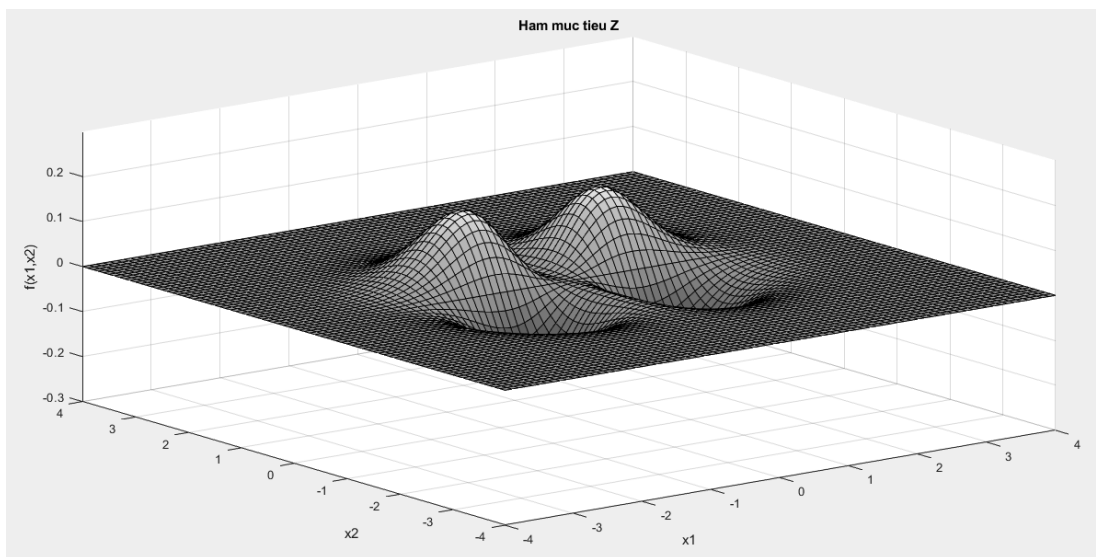
x = 1.0e+04 *-1.2479 -0.5022 ; fval = -6.0527e+20

Ví dụ khác:

Để tìm giá trị cực tiểu của hàm $Z = X^2 Y e^{-X^2 - Y^2}$ ta thực hiện chòm lệnh sau:

```
>> x = -4:0.1:4; y = -4:0.1:4;
[X,Y] = meshgrid (x, y);
Z = X.*X.*Y.*exp(-X.^2 - Y.^2);
>> figure
>> surf (X,Y,Z);
>> axis ([-4,4,-4,4,-0.3,0.3])
>> title ('Ham muc tieu Z')
>> xlabel ('x1')
>> ylabel ('x2')
>> zlabel ('f(x1,x2)')
```

Đồ thị của hàm được thể hiện ở Hình 2.



Hình 2: Đồ thị của hàm $Z = X^2 Y e^{-X^2 - Y^2}$

Ta soạn một script có tên function10.m với nội dung như sau và lưu vào một thư mục đã được khai báo ở MATLAB:

```
% function10.m
```

```
function F = function10(x)
F = x(1).*x(1).*x(2).*exp(-x(1).^2 - x(2).^2);
Vị trí cực tiểu đầu tiên nằm ở lân cận x = [2 -2].
Điểm [x1 x2] = [2 -2].
Ta xác định vị trí cực tiểu đầu tiên bằng các lệnh sau:
options = optimset('fminunc');
>> options = optimset(options, 'Display', 'iter', 'LargeScale', 'off');
>> x0 = [2 -2];
>> [x, fval] = fminunc('function10', x0, options)
Ta có kết quả như sau:
```

Iteration	Func-count	First-order f(x)	Step-size	Optimality
0	3	-0.0026837		0.00939
1	24	-0.155624	132.585	0.0435
2	30	-0.156386	0.1	0.0425
3	39	-0.157776	0.0292501	0.000954
4	42	-0.157777	1	4.26e-05
5	45	-0.157777	1	3.78e-07

Local minimum found.

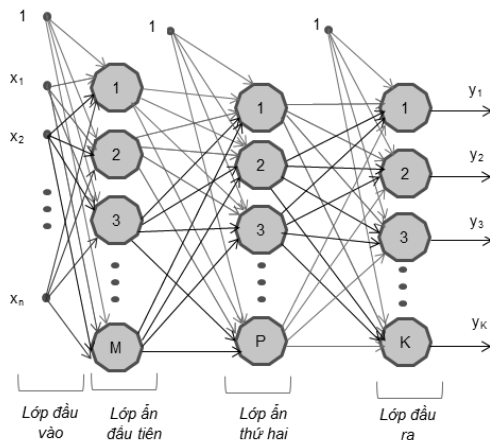
Optimization completed because the size of the gradient is less than the selected value of the optimality tolerance.

<stopping criteria details>

x = 1.0000 -0.7071

fval = -0.1578

2. Sử dụng mạng nơ ron MLP để ước lượng các thông số tài nguyên, môi trường



Hình 3: Sơ đồ mạng nơ ron MLP có hai lớp ẩn

Hình 3 mô tả sơ đồ mạng nơ ron MLP có hai lớp ẩn.

Theo [1] ta có thể xây dựng mạng MLP với số lượng lớp ẩn tùy ý, tuy nhiên các nghiên cứu đã chứng minh rằng, chỉ cần sử dụng tối đa hai lớp ẩn là có thể mô hình hóa một hàm phi tuyến với độ chính xác tùy chọn.

Hàm truyền đạt của đầu ra thứ i của mạng nơ ron MLP có một lớp ẩn, N đầu vào, M nơ ron trên lớp ẩn, K đầu ra là một hàm phi tuyến được thể hiện theo công thức sau [4]:

$$y_i = f_2 \left\{ \sum_{j=0}^M \left[f_1 \left(\sum_{k=0}^N x_k W_{jk} \right) V_{ij} \right] \right\}$$

Trong đó các thông số được mô tả ở Bảng 1.

Bảng 1. Mô tả các thông số của mạng nơ ron có một lớp ẩn

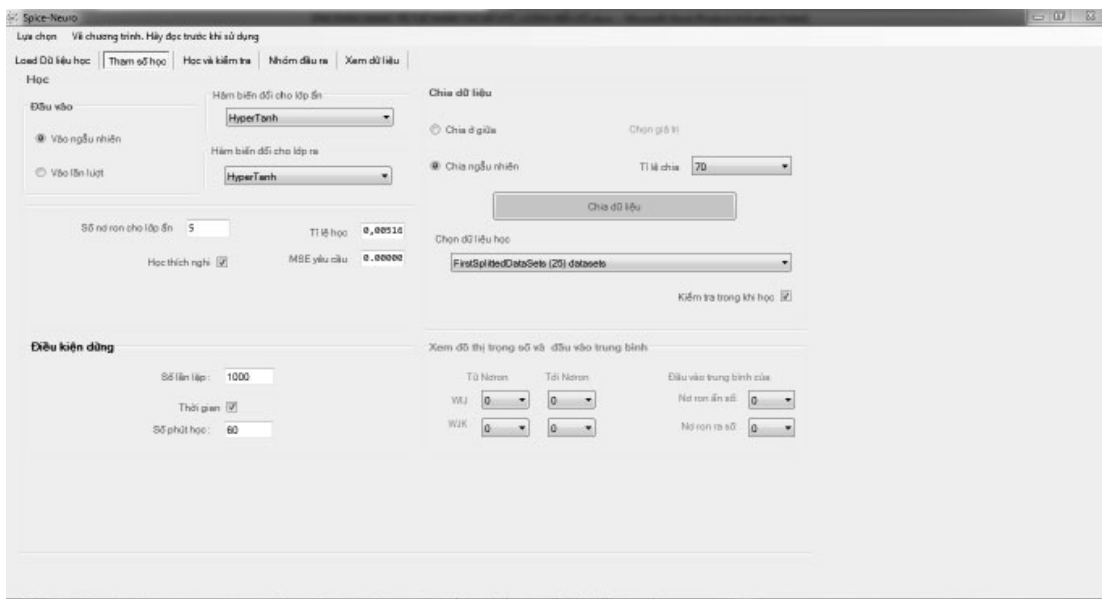
Ký hiệu thông số	Ý nghĩa	Ghi chú
g_i	Tổng các kích thích đầu vào của nơ ron đầu ra thứ i	$i = 1 \div K$ $g_i = \sum_{j=0}^M v_j V_{ij} g_i = \sum_{j=0}^M v_j V_{ij}$
K	Số đầu ra của mạng MLP một lớp ẩn	
M	Số nơ ron trên lớp ẩn	
N	Số đầu vào của mạng MLP một lớp ẩn	
u_j	Tổng các kích thích đầu vào của nơ ron thứ j của lớp ẩn ($j = 1 \div M$)	$u_j = \sum_{k=0}^N x_k W_{jk}$
V_{ij}	Trọng số ghép nối giữa lớp ẩn và lớp đầu ra	$i = 1 \div K$ $j = 1 \div M$
v_j	Đầu ra của nơ ron thứ j của lớp ẩn $j = 1 \div M$	$v_j = f_1(u_j)$ Đầu vào phân cực cho các nơ ron lớp ra được xem cố định ($v_0 = 1$)
x_k	Tín hiệu đầu vào thứ k với $k = 0 \div N$	Đầu vào phân cực cố định $x_0 = 1$
y_i	Tín hiệu đầu ra thứ i của mạng nơ ron với $i = 1 \div K$	$y_i = f_2(g_i)$
W_{jk}	Trọng số ghép nối giữa lớp đầu vào và lớp ẩn	$j = 1 \div M$ $k = 0 \div N$

Trên thực tế nhiều khi ta cần biểu đồ theo thời gian của một vài thông số tài nguyên, môi trường theo thời gian, chẳng hạn như vận tốc gió của một địa điểm, lưu lượng nước của một dòng sông,... Để có được thông số theo yêu cầu, ta phải tiến hành đo đạc trong một thời gian dài, có khi đến vài năm để có thể đưa ra một mô hình thống kê gần đúng. Thời gian đo dài sẽ làm tăng chi phí đồng thời làm chậm tiến độ triển khai dự án. Để giải quyết vấn đề trên, bài báo này đề xuất giải pháp rút ngắn thời gian đo đạc dựa trên các số liệu đo đạc quan trắc tại một điểm lân cận với vị trí đã chọn hoặc một vị trí khác có số liệu theo một quy luật tương quan nào đó. Trên thực tế của ngành tài nguyên môi trường, trong

mỗi vùng ta thường có các điểm thu thập dữ liệu quan trắc liên tục trong một thời gian dài, có thể tới hàng chục năm. Khi cần xác định thông số $Y_D(t)$ tại điểm D (viết tắt của điểm cần thực hiện Dự án), trước tiên ta tìm điểm lân cận gần đó là điểm T (T là viết tắt của điểm tham chiếu) mà tại đó dữ liệu $Y_T(t)$ đã được đo đạc nhiều năm. Nếu ta xây dựng được hàm quan hệ phụ thuộc giữa thông số của hai điểm thì dễ dàng ước lượng được thông số tại điểm D trong toàn bộ khoảng thời gian có dữ liệu đo đạc tại điểm T. Để tìm được mối tương quan này ta sẽ tiến hành đo đạc thông số tại điểm D trong một khoảng thời gian ngắn. Khoảng thời gian cụ thể này tùy thuộc vào yêu cầu của từng loại thông số đối với một dự án.

Trong thời gian này, tại điểm tham chiếu T vẫn tiếp tục có các dữ liệu của thông số. Từ những dữ liệu này ta sẽ xây dựng một hàm phụ thuộc giữa các thông số của điểm D và T. Hàm phụ thuộc này sẽ được mở rộng đối với toàn bộ thời gian đo đạc tại điểm D. Như vậy, ta sẽ có kết quả gần đúng trong một khoảng thời gian dài cho thông số tại điểm cần tiến hành dự án. Nói chung kết quả thực tế cho thấy các hàm phụ thuộc này là các hàm phi tuyến bậc cao. Mặt khác, các dữ liệu đo đạc chứa nhiều nhiễu nên ta thường lựa chọn mô hình xấp xỉ bằng các mạng nơ ron để đảm bảo hiệu quả tính toán với sai số cho phép. Bài báo này đề xuất chọn mô hình xấp xỉ bằng mạng MLP để tính toán cho các bài toán liên quan các tham số về tài nguyên, môi trường. Quá trình tối ưu hóa sai số được thực hiện bằng cách điều chỉnh thích nghi các trọng số ghép nối.

Dữ liệu vào thể hiện mối quan hệ của hai tham số tài nguyên, môi trường tương ứng với 2 địa điểm D và T được mô phỏng gồm 36 giá trị. Trên thực tế ta có thể mở rộng với số lượng nhiều hàng hơn. Ta chuyển dữ liệu thành file có dạng file text file TXT (Tab Separated Value File Format). Dữ liệu này được nhập vào phần mềm mạng nơ ron PLM 3 lớp [2]. Dữ liệu có thể được chia làm hai phần, một phần để học và một phần để kiểm tra. Ở đây việc chia dữ liệu ngẫu nhiên thành hai phần 70% và 30%, dùng 70% để học và 30% để kiểm tra. Tiếp theo, ta cần chọn số nơ ron cho lớp ẩn (hidden layer), số lần lặp, thời gian học và MSE (Mean of Square Error) yêu cầu. Ngoài ra ta có thể lựa chọn phương thức học thích nghi. Hình 4 biểu thị cửa sổ chọn các tham số học thích nghi.

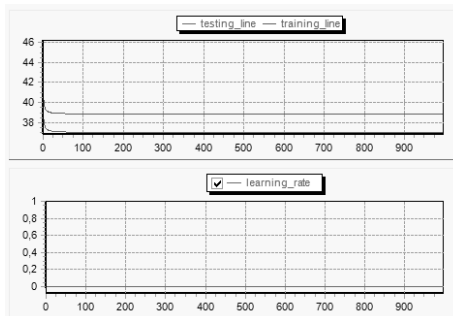


Hình 4: Cửa sổ chọn các tham số học thích nghi

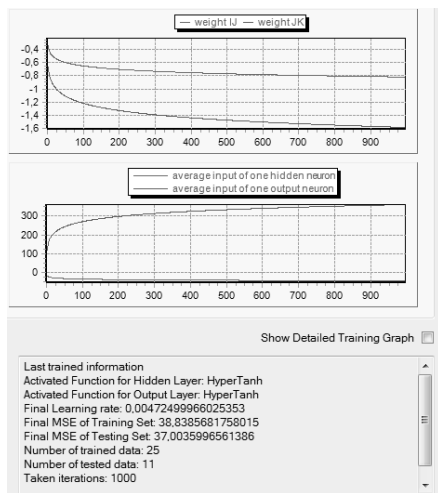
Nghiên cứu

3. Kết quả mô phỏng

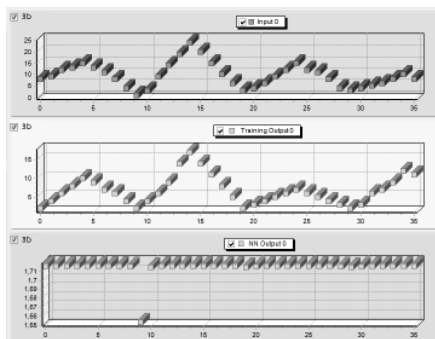
Sau khi đào tạo mạng, ta có thể kiểm tra lỗi trong quá trình học. Hình 5 thể hiện đồ thị tham số lỗi và tỷ lệ học.



Hình 5: Tham số lỗi trong quá trình học và tỷ lệ học



Hình 6: Đồ thị trọng số IJ, trọng số JK, đầu vào trung bình của một nơ ron ẩn, đầu vào trung bình của một nơ ron ra theo số bước học

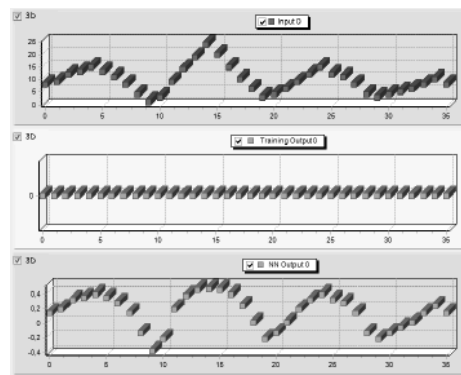


Hình 7: Tín hiệu đầu vào, tín hiệu đầu ra và tín hiệu đào tạo của mạng nơ ron

Hình 6 mô tả đồ thị trọng số IJ, trọng số JK, đầu vào trung bình của một nơ ron

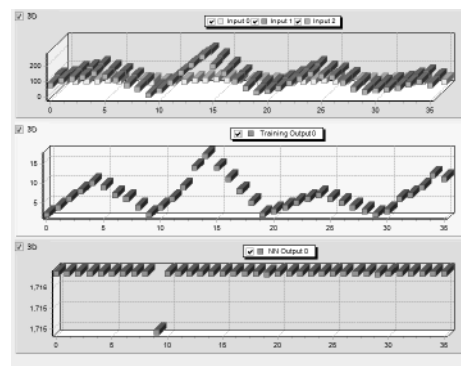
ẩn, đầu vào trung bình của một nơ ron ra theo số bước học. Hình 7 biểu diễn tín hiệu đầu vào, tín hiệu đầu ra và tín hiệu đào tạo của mạng nơ ron.

Ta có thể đưa dữ liệu mới hoặc dữ liệu cũ của vị trí tham chiếu T để suy ra dữ liệu cho vị trí dự án D. Như vậy khi ta nhập dữ liệu đầu vào thì với các trọng số vốn có do mạng đã học sẽ cho ta dữ liệu đầu ra mong muốn với sai số cho phép. Chẳng hạn ta sẽ có kết quả đối với 36 hàng dữ liệu đầu vào. Ví dụ, ta đưa file mới với dữ liệu đầu vào thành đầu vào của mạng nơ ron ta sẽ có kết quả đầu ra của mạng được biểu thị ở Hình 8.



Hình 8: Tín hiệu ra của mạng nơ ron tương ứng với tín hiệu vào đối với tập liệu mới

Ta có thể mở rộng bài toán cho nhiều đầu vào, nhiều đầu ra. Ví dụ dữ liệu vào tương ứng 3 đầu vào, một đầu ra được thể hiện như Hình 9.



Hình 9: Tín hiệu ra của mạng nơ ron tương ứng với 3 tín hiệu vào

4. Kết luận

Bài báo đã giới thiệu bài toán tìm giá trị cực trị đa biến thông qua ví dụ mô phỏng; đề xuất sử dụng mạng nơ ron MLP để ước lượng các thông số; đề xuất chọn mô hình xấp xỉ bằng mạng MLP để tính toán cho các bài toán liên quan các tham số về tài nguyên, môi trường. Quá trình tối ưu hóa sai số được thực hiện bằng cách điều chỉnh thích nghi các trọng số ghép nối. Ta có thể sử dụng mạng trí tuệ nhân tạo để ước lượng biểu đồ các thông số tài nguyên, môi trường theo thời gian đối với các địa điểm có dữ liệu tương quan. Như vậy, thời gian và kinh phí đo đạc dữ liệu được giảm, mà vẫn đảm bảo sự vận hành của các công trình với một sai số cho phép. Phương pháp này có thể dùng để xác định tốc độ gió tại một địa điểm căn cứ dữ liệu đo đạc đầy đủ hơn của một điểm lân cận

trong một khoảng thời gian dài. Ngoài ra phương pháp này có thể được dùng để dự báo lưu lượng nước của lòng hồ, con sông hoặc dự báo mức đỉnh và đáy của các hồ, con sông trong các đợt mưa lũ.

TÀI LIỆU THAM KHẢO

- [1]. Boger B, Guyon I (1997). *Knowledge extraction from artificial neural network models*. IEEE Systems, Man, and Cybernetics Conference.
- [2]. Cao Thang (2007). *Intructions for using Spice-MLP software*. Soft Intelligence Laboratory, Ritsumeikan University, Japan.
- [3]. Nguyễn Phùng Quang (2008). *Matlab & Simulink dành cho kỹ sư điều khiển tự động*. Nhà xuất bản Khoa học và Kỹ thuật.
- [4]. PGS. TSKH Trần Hoài Linh (2019). *Mạng nơ ron và ứng dụng trong xử lý tín hiệu số*. Nhà xuất bản Bách Khoa, Hà Nội.

BBT nhận bài: 25/5/2020; Phản biện xong: 10/6/2020; Chấp nhận đăng: 26/6/2020