

MẠNG NƠON NHIỀU LỚP LAN TRUYỀN NGƯỢC DÙNG CHO VIỆC MÔ HÌNH HÓA

THE BACKPROPAGATION NEURAL NETWORK FOR MODELLING

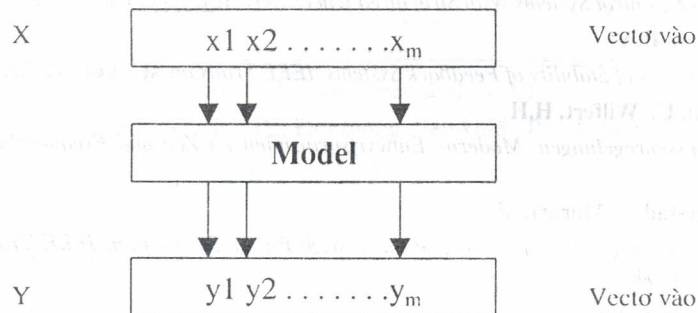
TRẦN NGỌC HÀ¹, NGUYỄN THANH THỦY²

Abstract. In this paper we shall investigate some aspects on using the back propagation neural network for modeling. As an illustration, an experimental application to modelling equilibrium data of the liquid-liquid system DyOHCIODEHPA has been taken. It is shown that the neural network approach is an effective method for modelling various systems of this kind. Experimental results showed that errors of prediction of the proposed neural network are really in the experimental range.

1. ĐẶT VẤN ĐỀ:

Mô hình hóa là giải pháp ứng dụng toán học được áp dụng rộng rãi trong các ngành khoa học tự nhiên như vật lý, hóa học, dược, thực phẩm v.v... Thực tế cho thấy nó thực sự hữu hiệu trong việc nghiên cứu các công thức tối ưu để điều chế, sản xuất các sản phẩm trong công nghiệp [1,2].

Về thực chất, mô hình hóa là quá trình tìm kiếm hàm giải tích hoặc một thủ tục cho phép sản sinh các vectơ lời giải đầu ra (output) Y gồm n biến tương ứng với mỗi vectơ đầu vào (input) X gồm m biến.



Hình 1: Mô hình sản sinh vectơ ra Y từ vectơ vào X

Các mô hình dạng này thường thỏa mãn điều kiện là: Đối với sự thay đổi nhỏ của vectơ vào chỉ tạo ra sự thay đổi nhỏ của vectơ ra. Việc xác định mô hình dựa trên một tập hữu hạn các số liệu thực nghiệm $\{X_i, Y_i\}$. Các số liệu thực nghiệm này được đo theo một kế hoạch hóa thực nghiệm nhất định.

Các kỹ thuật mô hình hoá thông thường cho một quá trình đòi hỏi các dạng hàm toán học phải được biết trước hoặc sinh ra từ lý thuyết của quá trình hoặc chỉ đoán theo kinh nghiệm. Đối với mỗi biến của vectơ ra Y_i phải xây dựng một hàm f_i .

$$y_1 = f_1(x_1, \dots, x_m, a_{11}, \dots, a_{1p})$$

$$y_2 = f_2(x_1, \dots, x_m, a_{21}, \dots, a_{2r})$$

¹ Viện Công nghệ Xạ hiếm, Hà Nội

² Đại học Bách Khoa, Hà Nội.

MẠNG NƠN NHIỀU LỚP LAN TRUYỀN NGƯỢC DÙNG CHO VIỆC MÔ HÌNH HÓA

$$y_n = f_n(x_1, \dots, x_m, a_{n1}, \dots, a_{ns})$$

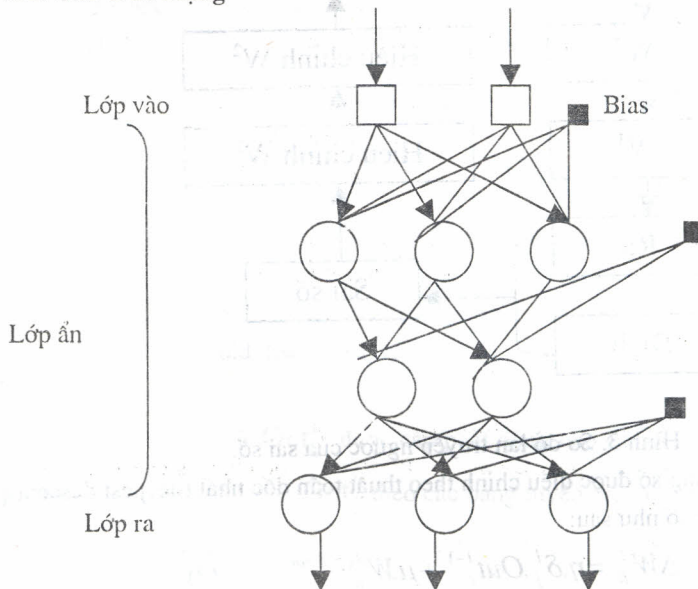
Trong đó a_{ij} là các tham số chưa biết của mô hình.

Khi xây dựng mô hình phải chỉ rõ số lượng các tham số cũng như vị trí của chúng trong các hàm f_i . Việc xác định mô hình là việc xác định tất cả các tham số a_{ij} dựa trên sự phù hợp giữa giá trị tính toán và giá trị thực nghiệm cho cùng một vector vào. Phương pháp bình phương tối thiểu thường được dùng cho quá trình xác định này. Mô hình sẽ dựa báo tốt nhất khi các số liệu thực nghiệm được phân bố đều trong toàn bộ không gian biến số đầu vào với mật độ thỏa đáng [1]. Dễ nhận thấy đối với các quá trình phức tạp khi số lượng biến đầu vào và đầu ra tăng lên việc tìm kiếm, xác định một mô hình như vậy hoàn toàn không đơn giản.

Mạng nơron nhiều lớp lan truyền ngược là một giải pháp hữu hiệu cho việc mô hình hóa, đặc biệt đối với các quá trình phức tạp hoặc cơ chế chưa được biết rõ ràng. Nó không đòi hỏi bất kỳ sự hiểu biết trước về các dạng hàm cũng như các tham số. Thực tế nó hoạt động như một hộp đen. Với một số đủ lớn các trọng số (weight) mạng hoàn toàn có đủ độ tự do để biểu diễn mối quan hệ giữa các đại lượng vào và ra. Mạng sử dụng tập số liệu thực nghiệm cho quá trình học hay quá trình điều chỉnh các trọng số. Việc xác định mô hình đơn giản là việc lựa chọn một cấu trúc mạng và cho mạng học số liệu. Các thông số sau quá trình học dùng để đánh giá chất lượng của mạng. Quá trình học có thể lâu song một khi đã ổn định, việc dựa báo trở nên rất nhanh.

2. MÔ HÌNH MẠNG NƠN NHIỀU LỚP LAN TRUYỀN NGƯỢC

2.1. Cấu trúc mạng



Hình 2. Cấu trúc mạng lan truyền ngược với 2 lớp ẩn.

Mạng nơron lan truyền ngược luôn có một lớp vào, một lớp ra và không có hoặc có các lớp ẩn. Các nơron trên các lớp được xếp tuyến tính. Số lượng nơron của lớp vào bằng số biến của vector vào số nơron của lớp ra bằng biến số của vector ra. Các lớp của mạng được nối với nhau theo cách đầy đủ [1]. Hình 2 là một ví dụ về cấu trúc của một mạng có hai lớp ẩn. Ngoài ra trên các lớp trừ lớp ra có thêm một tham số điều khiển (bias). Số lớp cũng như số lượng các nơron trên mỗi lớp tùy thuộc vào từng ứng dụng. Trên thực tế việc xây dựng mạng dựa trên việc thử và sai (trial and errors). Các mạng được sử dụng thường chỉ gồm 3 lớp: lớp vào, lớp ra và một lớp ẩn.

2.2. Thủ tục lan truyền:

Giả sử mạng có L lớp, và trên lớp thứ l có N_l nơron. Thủ tục lan truyền của mạng như sau:

$$Out_j^l = F\left(\sum_{i=1}^{N_{l-1}+1} W_{ji}^l \cdot Out_i^{l-1}\right) \quad (1)$$

Trong đó:

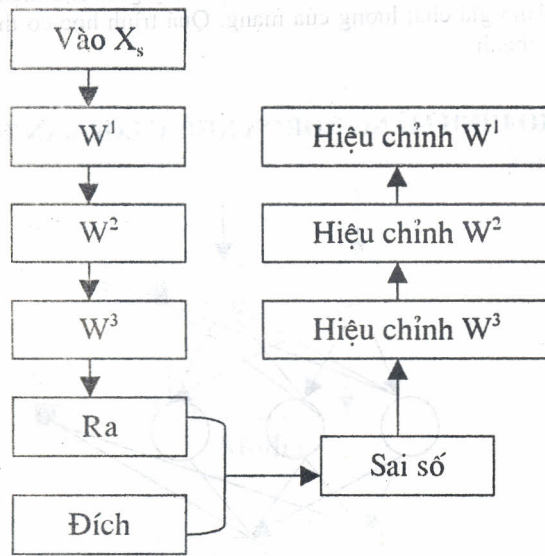
W_{ji}^l - Trọng số của đầu vào thứ i của neuron thứ j trên lớp l .

Out_j^{l-1} - đầu ra của neuron thứ j trên lớp $l-1$ với $Out^0(x_1, \dots, x_m, 1)$.

Thông thường F là hàm biến đổi sigmoidal có dạng $F(x) = \frac{1}{1 + e^{-x}}$

2.3. Thủ tục học

Phương pháp lan truyền ngược (hình 3) là phương pháp học có giám sát (supervised) [1]. Do đó nó cần một tập các đôi vector vào và đích (X_s, Y_s).



Hình 3. Sơ đồ lan truyền ngược của sai số.

Trong quá trình học các trọng số được điều chỉnh theo thuật toán dốc nhất (deepest descent) [1].

Công thức điều chỉnh trọng số như sau:

$$\Delta W_{ji}^l = \eta \cdot \delta_j^l \cdot Out_i^{l-1} + \mu \cdot W_{ji}^{l(previous)} \quad (2)$$

$$W_{ji}^{l(new)} = W_{ji}^{l(old)} + \Delta W_{ji}^l$$

Trong đó:

η - hằng số học

μ - hằng số quán tính

$$\delta_j^l = (y_j - Out_j^l) \cdot Out_j^l \cdot (1 - Out_j^l) \text{ đối với lớp ra}$$

$$= \left(\sum_{k=1}^{N_{l+1}} \delta_k^{l+1} \cdot W_{kj}^{l+1}\right) \cdot (1 - Out_j^l) \text{ đối với các lớp ẩn}$$

Thủ tục học gồm các bước sau:

- Lấy lần lượt các cặp số liệu $X(x_1, \dots, x_m)$, $Y(y_1, \dots, y_n)$ từ tệp số liệu.
- Xác định vectơ Out $(x_1, \dots, x_m, 1)$
- Lan truyền vectơ Out qua mạng theo công thức (1)
- Điều chỉnh các trọng số theo công thức (2).
- Thực hiện lại các bước trên với cặp (X, Y) tiếp theo.

Mỗi lượt cho phép tất cả các số liệu đã truyền qua mạng được gọi là một lần lặp. Sau mỗi lần lặp, tham số RMS (Root mean square error) được tính:

$$RMS = \sqrt{\frac{\sum_{s=1}^p \sum_{i=1}^n (y_i - \text{Out}_i^s)^2}{p \cdot n}} \quad (3)$$

Trong đó p là số bộ số liệu. Thông thường cần hàng nghìn thậm chí chục nghìn lần lặp để mạng hội tụ. Khả năng, tốc độ hội tụ của mạng phụ thuộc vào kích thước và các tham số học của mạng [1].

3. CÀI ĐẶT MẠNG NƠ RON LAN TRUYỀN NGƯỢC.

Một mạng tổng quát được cài đặt trên C như một lớp (class) Network. Các tham số của mạng là các biến thành viên bảo vệ còn các chức năng của mạng được thiết kế cho các hàm thành viên của lớp.

```
class Network
{ protected:
    int L, *N;
    float η, μ *W, *Out;
    void lan_truyền();
public:
    Network(in, int*);
    ~Network();
    void học(char* str, int option);
    void đoán(float* X, float* Y);
};
```

Các trọng số W_{ji}^l và Out_j^l được cấp phát động cho các con trỏ W và Out tương ứng. Việc tính địa chỉ k cho các phân tử W_{ji}^l và Out_j^l theo các bảng tuyến tính W và Out được thực hiện bởi các thủ tục như sau:

```
/* tính địa chỉ k cho */
{k = (j-1)*(Nl-1+1)+i-1'
if (l>1) for (s=1; s<l; s++) k += Ns*(Ns-1+1);
}
/*tính địa chỉ k cho Outjl*/
{k=1+j-1;
if (l>0) for (s=0; s<l; s++) k += Ns;
}
```

Thủ tục lan truyền và học:

```
{*Lan truyền* }
{for (l=1; l<L; l++)
```

```

for(j=1;j<Ni;j++)
{for(Net=0,i=1;i<=Nl-1+1;i++)Net+= Wjil * Outjl;
Outjl =F*(Net);
}
}
/*Học*/
{lan truyền();
for(l=L-1;l>0;l--)
for(j=1;j<Nl;j++)
{s0= Outjl*(1-Outjl);
if(l==L-1) s=yj- Outjl;
else for(s=0, k=1, k<=Ni+1; k++) s+= δkl+1 * Wkjl+1;
δjl =s* s0;
for(i=1;i<Ni-1+1;i++); ΔWkjl+1 = η.δjl.Outil-1 + μ.ΔWjil
}
for(l=1;l<L;l++)
for(j=1; j<=Nl;l++)
for(i=1;i<Ni-1+1; i++) Wjil = ΔWjil
}

```

4. CÁC NGHIÊN CỨU VỀ MẠNG NƠN LAN TRUYỀN NGƯỢC

4.1. Tổ chức số liệu

Số liệu được tổ chức trong một tệp số liệu. Các cặp (X_n, Y_n) được viết lần lượt theo từng dòng. Do đặc điểm của hàm biến đổi F các số liệu này được định cỡ vào khoảng (0,1). Các phương pháp định cỡ tuyến tính cũng như không tuyến tính được thử nghiệm cho thấy định cỡ tuyến tính tốt hơn và đơn giản hơn. Khoảng định cỡ (0.2,0.8) cho thấy là tốt nhất. Trong khoảng đó hàm F gần như tuyến tính. Với khoảng nhỏ hơn như (0.3,0.7) không có cải thiện gì hơn nhưng lại gây sai số lớn hơn trong quá trình giải định cỡ từ đầu ra của mạng. Các sai số trong dự báo của mạng phụ thuộc rất nhiều vào sai số của số liệu. Do đó số liệu cho mạng học nên là các số liệu được đo trực tiếp. Không nên sử dụng các số liệu dẫn xuất từ các số liệu đo.

Trong quá trình học, tập số liệu được chia làm hai phần: phần học và phần kiểm tra. Phần kiểm tra được chọn ngẫu nhiên theo một xác suất p. Tùy thuộc vào số lượng số liệu giá trị p có thể chọn trong khoảng (0.0,0.6). Một đại lượng $RMS_M=(1-p).RMS_H+p.RMS_T$ được tính cho mỗi lần lặp. Trong đó RMS_H là RMS của phần học và RMS_T cho phần kiểm tra.

Khi sai số tương đối quan trọng hơn sai số tuyệt đối việc tính sai số của lớp ra trong (2) được tính theo công thức mới như sau:

$$\delta_j^l = (y_j - Out_j^l) \cdot Out_j^l \cdot (1 - Out_j^l) / y_j$$

4.2. Cấu trúc và tham số học của mạng:

Việc xây dựng cấu trúc mạng (số lớp, số nơon) là công việc mang tính thử nghiệm. Tuy nhiên với số lượng các trong số nhỏ, mạng thể không có đủ độ tự do. Trong khi đó nếu số trọng số quá lớn, rất có thể dẫn đến hiện tượng bão hòa (overtraining effect). Đó là hiện tượng mạng đã rất thuộc các số liệu học,

song dự báo lại cho sai số lớn. Các nghiên cứu cho thấy việc lựa chọn tốt đạt được khi số trọng số nhỏ hơn hoặc gần bằng số bộ số liệu khác nhau để học. Do đó việc thử nghiệm thường xuất phát từ các mạng có kích thước nhỏ.

Giá trị của các trọng số ban đầu không thật sự quan trọng lắm đối với quá trình học nhưng phải được gán bằng các số ngẫu nhiên tuy nhỏ nhưng không đồng đều bằng 0. Nếu các giá trị này lớn, mạng không thể có khả năng học. Điều này là rõ ràng vì khi với các đối số có giá trị tuyệt đối lớn, hàm F cho kết quả là hai giá trị 0 và 1. Các trọng số ban đầu được gán giá trị trong khoảng $(-1/n, 1/n)$ (n- số trọng số trên từng lớp) sẽ cho kết quả tốt. Các hàm sinh số ngẫu nhiên trong các máy PC đảm nhiệm tốt quá trình này.

Nhân tố quan trọng nhất là lựa chọn các hằng số học η và hằng số quán tính μ . Tổng của hai hằng số này dao động xung quanh giá trị $1(\eta + \mu \approx 1)$. Với giá trị $\eta = \mu$ có thể dẫn đến sự dao động trong quá trình học. Khi hằng số học quá lớn, quá trình học sẽ nhanh, song rất có thể kết thúc sớm, khi rơi vào một cực tiểu cục bộ. Hằng số quán tính lớn làm cho quá trình học chậm lại, nhưng giúp cho mạng thoát ra khỏi các cực trị cục bộ. Từ những tính chất trên cho thấy tốt nhất là khi giá trị μ lớn hơn không nhiều giá trị η . Các giá trị này thay đổi tùy từng ứng dụng, thực tế cho thấy giá trị $\mu = 0.55$ và $\eta = 0.5$ khá tốt cho sự lựa chọn ban đầu. Mạng được chọn làm mô hình sẽ là mạng có giá trị RMS_M nhỏ nhất.

5. MÔ HÌNH HÓA SỐ LIỆU CÂN BẰNG CHIẾT LỎNG - LỎNG

5.1. Đặc điểm hệ chiết lỏng - lỏng:

Kỹ thuật chiết lỏng - lỏng là kỹ thuật được sử dụng nhiều trong hóa học để tách, tinh chế các hợp chất hóa học. Trong trường hợp tổng quát các giá trị đo được bằng thực nghiệm là nồng độ các cấu tử ban đầu C, giá trị axit ban đầu H, giá trị nồng độ cân bằng ở pha nước C_{aq} và nồng độ C_{or} trên pha hữu cơ. Xác định các hàm.

$$D = \frac{C_{or}}{C_{aq}} = f(C, H) \text{ hoặc } C_{aq} = f(C, H)$$

rất quan trọng cho việc xác định các điều kiện tối ưu cho hệ chiết liên tục. Việc xây dựng mô hình lý thuyết gặp nhiều khó khăn nên đã có một mô hình thực nghiệm được đề nghị [2].

$$D = A_1 C^{A_2} \exp(-A_3 C^{A_4} H)$$

Trong đó A_1, A_2, A_3, A_4 là các tham số của mô hình. Với mô hình này sai số tương đối trung bình cho hệ chiết Dy-HCl-DEHPA (bi 2-ethylhexylphosphoric acid) lên tới 10-12% [3].

5.2. Mô hình hóa bằng mạng nơon:

Trong số 49 số liệu thực nghiệm ở bảng 1[3], 5 giá trị (**) đã được giành riêng cho quá trình dự báo. Mạng nơon xây dựng gồm 3 lớp từ kích thước nhỏ ($2 \times 2 \times 1$) đến kích thước lớn ($2 \times 7 \times 1$). Giá trị xác suất kiểm tra trong quá trình học là $p=0.2$, $\eta=0.5$, $\mu=0.55$. Thời gian học sau 10.000 lần lặp tăng lên theo kích thước mạng. Đối với mạng lớn nhất (2.7×1) là 22 phút trên máy PC 486-66 MHz. Bảng 2 thống kê sự phụ thuộc giá trị RMS_H và RMS_T vào kích thước mạng. Kết quả cho thấy ở các mạng lớn có xảy ra quá trình học quá. Mạng ($2 \times 7 \times 1$) là tối ưu cho hệ khảo sát. Ma trận trọng số của mạng ($2 \times 7 \times 1$) được liệt kê ở bảng. Các giá trị tính toán và dự báo so sánh với giá trị thực nghiệm được trình bày trong bảng 1. Sai số của quá trình dự báo là 4.52% nằm trong khoảng sai số thực nghiệm (3-4%)[3].

KẾT LUẬN.

Từ ví dụ minh họa mô hình hóa hệ chiết lỏng - lỏng cho thấy việc sử dụng mạng nơon nhiều lớp lan truyền ngược cho việc mô hình hóa đơn giản và hiệu quả hơn so với các phương pháp thông thường. Một trong những lợi thế chính là nó không đòi hỏi hiểu biết trước về các dạng hàm toán học. Lợi thế này xuất phát từ khả năng học mẫu số liệu của mạng nơon nói chung. Tuy nhiên với phương pháp học dốc nhất quá trình học luôn hội tụ là không đảm bảo. Một giải pháp học theo giải thuật di truyền có lẽ sẽ tốt hơn và cần được nghiên cứu.

Bảng 1. Các giá trị C, H và C_{aq}. Hệ chiết Dy-HCl-DEHPA

No	C	H	C _{aq} (TN)	C _{aq} (TT)	No	C	H	C _{aq} (TN)	C _{aq} (TT)
1.	0,10	0,371	0,031	0,038	26	0,30	0,916	0,244	0,239*
2.	0,10	0,501	0,038	0,045**	27	0,30	1,012	0,250	0,246
3.	0,10	0,531	0,045	0,046	28		1,099	0,256	0,252
4.	0,10	0,642	0,055	0,052	29	0,30	1,253	0,265	0,263
5.	0,10	0,693	0,063	0,055	30	0,40	-0,288	0,265	0,253
6.	0,10	0,811	0,069	0,061	31	0,40	0,000	0,276	0,274
7.	0,10	0,916	0,073	0,066	32	0,40	0,223	0,290	0,291
8.	0,10	1,099	0,078	0,076	33	0,40	0,405	0,301	0,304
9.	0,10	1,253	0,082	0,084	34	0,40	0,560	0,310	0,315
10.	0,20	0,288	0,078	0,075	35	0,40	0,693	0,320	0,325
11.	0,20	0,000	0,086	0,092	36	0,40	0,811	0,331	0,333
12.	0,20	0,223	0,096	0,105	37	0,40	0,916	0,339	0,341
13.	0,20	0,405	0,114	0,115**	38	0,40	1,012	0,345	0,348**
14.	0,20	0,693	0,138	0,133	39	0,40	1,099	0,351	0,354
15.	0,20	0,916	0,154	0,146	40	0,50	-0,288	0,365	0,357
16.	0,20	1,012	0,159	0,152	41	0,50	0,000	0,379	0,377
17.	0,20	1,099	0,163	0,158	42	0,50	0,223	0,391	0,392**
18.	0,20	1,253	0,169	0,167	43	0,50	0,405	0,400	0,404
19.	0,30	-0,288	0,162	0,157	44	0,50	0,560	0,407	0,414
20.	0,20	0,000	0,176	0,176	45	0,50	0,693	0,417	0,422
21.	0,30	0,223	0,191	0,191	46	0,50	0,811	0,427	0,429
22.	0,30	0,560	0,201	0,204	47	0,50	0,916	0,435	0,435
23.	0,30	0,560	0,211	0,214	48	0,50	1,012	0,442	0,440
24.	0,30	0,693	0,224	0,224	49	0,50	1,099	0,445	0,445
25.	0,30	0,811	0,234	0,232					

** giá trị được Mạng Noron đoán nhận.

Bảng 2. Sự phụ thuộc giá trị RMS_H(%) RMS_T và sai số đoán nhận P vào kích thước mạng sau 10.000 lần lặp.

Kiến trúc mạng	2x2x1	2x5x1	2x6x1	2x7x1	2x8x1	2x9x1	2x12x1	2x15x1	2x17x1
RMS-H	2,858	3,325	3,070	2,117	2,034	2,780	3,529	4,523	5,594
RMS-T	4,291	2,716	2,529	2,956	4,040	4,530	1,483	3,902	5,079
P									

Bảng 3. Ma trận trong số của mạng nơ ron (2x7x1) sau 10.000 lần lặp.

-3,200 -0,379 -0,799
 -2,567 -0,669 -0,795
 W= -3,226 -0,796 -0,295
 -1,661 -0,845 -1,110
 -2,128 -0,866 -0,701
 -4,336 -1,086 4,418
 -1,669 -0,852 -1,074
 -1,702 -1,541 -1,65 -1441 -1,398 -3,358 -1,413 3,439

TÀI LIỆU THAM KHẢO

1. J.Zupan, J. Gasteiger, *Neural network for chemists*. VCH.1993
2. T.Sekin, *Solvent extractiob*, Tokyo, 1990.
3. S.L. Miskra, *The extraction of Yttrium, Dysposium and Base metals using organophosphorus acids*, PH.D.Thesis, Bombay, 1995.