

ỨNG DỤNG NEURO-FUZZY VÀ KỸ THUẬT TRỰC QUAN HÓA KHẢO SÁT MỐI LIÊN QUAN NHÂN QUẢ TRONG CÔNG THỨC DƯỢC PHẨM

Nguyễn Đăng Khoa⁽¹⁾, Đỗ Quang Dương⁽²⁾

(1) Trường Cao Đẳng Nghề Đông Nai, Biên Hòa

(2) Trường Đại Học Y Dược Tp. Hồ Chí Minh

(Bài nhận ngày 01 tháng 09 năm 2009, hoàn chỉnh sửa chữa ngày 15 tháng 03 năm 2010)

TÓM TẮT: *Neuro-fuzzy system là hệ thống được xây dựng bằng cách kết hợp mạng neuron (Neural Network) và logic mờ (Fuzzy logic) nhằm rút ra quy luật nhân quả từ dữ liệu thực nghiệm. Bài báo này trình bày kết quả ứng dụng của hai kỹ thuật trên vào việc tìm quy luật nhân quả trong công thức viên nén phóng thích có kiểm soát và biểu diễn kết quả trực quan bằng đồ thị kết hợp song song (Parallel Coordinate Graph).*

1. GIỚI THIỆU

Giai đoạn nghiên cứu và phát triển thuốc có liên quan đến hai nội dung chính: thành lập công thức và xây dựng quy trình sản xuất, trong đó xây dựng công thức là nền tảng. Đối với nhà sản xuất thuốc, việc thành lập công thức là việc thường xuyên phải làm, bởi vì mỗi sản phẩm đều có một vòng đời nhất định và nhu cầu cạnh tranh trên thị trường đòi hỏi phải không ngừng cải tiến sản phẩm hiện có hoặc thay thế bằng sản phẩm mới [1].

Các phương pháp tối ưu hóa truyền thống (toán thống kê, đơn hình...) có thể áp dụng với các dữ liệu đơn giản và tuyến tính. Nhưng với những dữ liệu phức tạp và phi tuyến thì các phương pháp này không còn phù hợp. Ngoài ra, các phương pháp truyền thống không tối ưu hóa được đồng thời nhiều biến phụ thuộc trong khi mỗi sản phẩm thường có rất nhiều tính chất cần được tối ưu.

Việc xây dựng và tối ưu công thức bằng máy tính ngày càng được áp dụng phổ biến vì các lợi ích của nó như không giới hạn số biến độc lập, có thể tối ưu đồng thời nhiều biến phụ thuộc, phù hợp với các loại dữ liệu phức tạp và phi tuyến.

Có rất nhiều phương pháp xây dựng và tối ưu công thức bằng máy tính đã được áp dụng như: mạng neuron (Neural Network), logic mờ (Fuzzy Logic), thuật giải di truyền (Genetic Algorithms) và các hệ thống lai giữa chúng. Những kỹ thuật này phần nào giải quyết được yêu cầu tối ưu như giảm thời gian và chi phí nghiên cứu và phát triển công thức. Tuy nhiên, vẫn còn một số hạn chế như: Mạng neuron chỉ cho được kết quả mà không cho biết được quá

trình thực hiện. Logic mờ sinh ra được các luật ở dạng "If ... then ..." nhưng không có khả năng học. Kết hợp những kỹ thuật này ta có thể giúp ta tìm được quy luật nhân quả trong công thức thuốc và tối ưu bằng giải thuật di truyền [5][6].

Bài báo này giới thiệu sự kết hợp giữa mạng neuron và logic mờ để tìm hiểu quy luật nhân quả trong công thức dược phẩm đồng thời thể hiện trực quan bằng đồ thị kết hợp song song (Parallel Coordinate Graph) giúp người sử dụng có thể hiểu rõ hơn các tri thức rút ra được từ hệ thống.

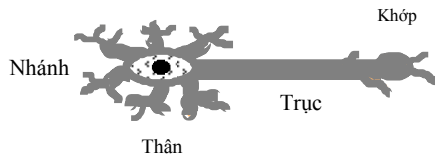
Mô hình này còn có thể được phát triển trong các lĩnh vực khác như chế biến mỹ phẩm, thuốc trừ sâu, và các sản phẩm công nghiệp khác,...

2. KỸ THUẬT NEURO-FUZZY VÀ KỸ THUẬT TRỰC QUAN HÓA

2.1. Mạng neuron

Mạng neuron (Neural networks) [5][6] đã được áp dụng từ hơn 60 năm qua, có ích trong việc thiết lập mô hình quan hệ nhân quả, đặc biệt đối với dữ liệu phi tuyến hay dữ liệu phức tạp.

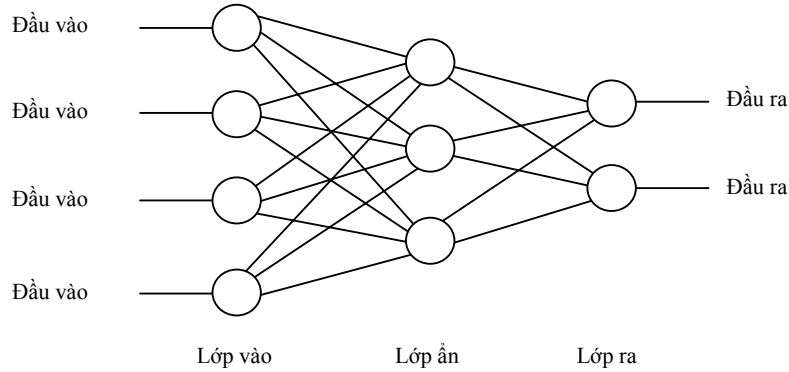
Mạng neuron sinh học được tạo thành bởi sự liên kết giữa rất nhiều tế bào thần kinh (neuron). Các xung động thần kinh được truyền từ khớp qua trục để đến thân. Thân tế bào tổng hợp các xung động thần kinh và ra quyết định tiếp tục truyền các tín hiệu sang tế bào thần kinh khác.



Hình 1. Cấu trúc của một tế bào thần kinh sinh học.

Mạng neuron nhân tạo là sự mô phỏng cấu trúc của mạng neuron sinh học, được tạo thành bởi sự liên kết giữa rất nhiều đơn vị thần kinh (perceptron). Những đơn vị thần kinh có nhiệm vụ thu thập các tín hiệu, xác định trọng số w_i , tổng cộng và chuyển các tín hiệu ấy sang các đơn vị thần kinh khác.

Cấu trúc của một mạng neuron thông thường được minh họa như sau:



Hình 2. Cấu trúc của một mạng neuron thường gặp

2.2 Neuro-Fuzzy

Neuro-fuzzy systems [4][6][7] là một hệ suy diễn mờ được tăng cường thêm khả năng học của mạng neuron. Trong hệ thống này, mạng neuron được đưa vào làm tăng khả năng tự điều chỉnh các hệ số (biến) trong các luật mờ và hàm thành viên của hệ thống. Với sự kết hợp này, khả năng học và suy diễn của hệ thống sẽ tốt hơn so với mạng neuron thông thường và tốc độ học cũng nhanh hơn. Các dạng NFS đã được giới thiệu: GARIC, FALCON, ANFIS, NEFCON, FUN, SONFIN, FINEST, EFuNN, dmEFuNN.

Mamdani Neuro-fuzzy System (Hình 3)

Lớp 1 (*input layer*): Là lớp nhập, chứa các nút nhập, các giá trị này truyền trực tiếp vào lớp 1.

Lớp 2 (*fuzzification layer*): Là lớp mờ hóa, mỗi nút i trong lớp này được gắn với một hàm thành viên $\mu(x)$, với x là đầu vào trực tiếp từ lớp 1. Giá trị xuất của mỗi lớp trong mức này là giá trị mờ của nút nhập được tính thông qua hàm thành viên $\mu(x)$.

Lớp 3 (*rule antecedent layer*): Là lớp tiền đề của luật, mỗi nút trong lớp này tương ứng với một tiền đề của một luật. Thông thường toán tử T-norm được sử dụng trong lớp này.

Lớp 4 (*rule consequent layer*): Là lớp kết luận của luật, mỗi nút trong lớp này kết hợp các tiền đề của luật lại và quyết định giá trị các nhân đầu ra (như High, Medium, Low,...). Số nút trong lớp này chính là số luật thu được.

Lớp 5 (*combination and defuzzification layer*): Là lớp kết hợp các luật và giải mờ, lớp này dùng toán tử T-norm để kết hợp các luật và tính toán các giá trị thô sau khi giải mờ.

Takagi-Sugeno Neuro-fuzzy System (Hình 4)

Lớp 1, 2, 3 có chức năng giống như mô hình Mamdani.

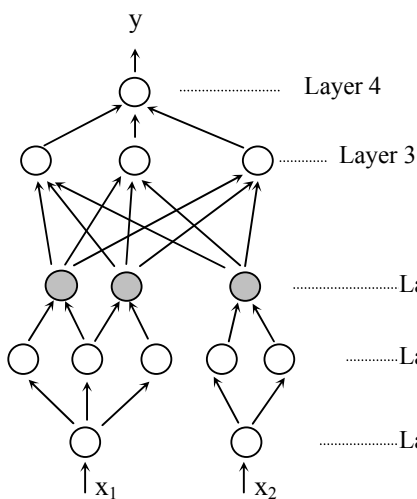
Lớp 4 (*rule strength normalization*): mỗi nút trong lớp này sẽ nhận giá trị xuất của lớp trước và sau đó tính tỷ số như sau:

$$\bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2, \dots$$

Lớp 5 (*rule consequent layer*): Là lớp kết luận của luật, mỗi nút ở lớp này xem như là nút điều chỉnh với kết quả xuất ở nút cuối, giá trị xuất ở mỗi nút được xác định theo công thức sau:

$$\bar{w}_i f_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i)$$

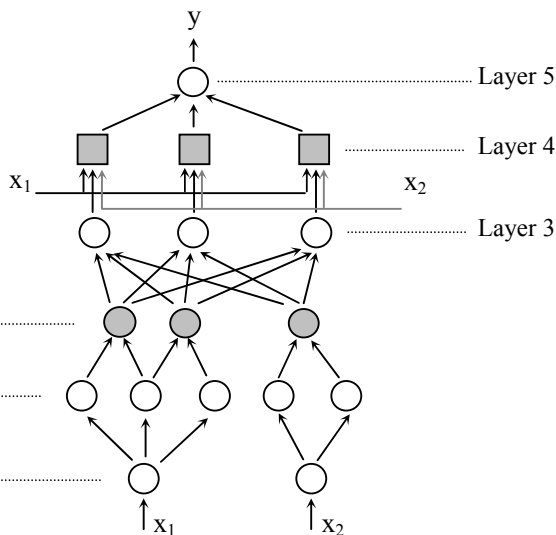
Với \bar{w}_i lấy từ lớp 4, và $\{p_i, q_i, r_i\}$ là các tham số tổng hợp.



Hình 3. Mamdani Neuro-fuzzy system

Lớp 6 (*rule inference layer*): Ở lớp này chỉ có một nút tổng hợp kết quả xuất ra từ lớp trước, giá trị xuất ở nút này là tổng các giá trị kết quả xuất ra từ lớp trước như công thức:

$$Output = \sum \bar{w}_i f = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

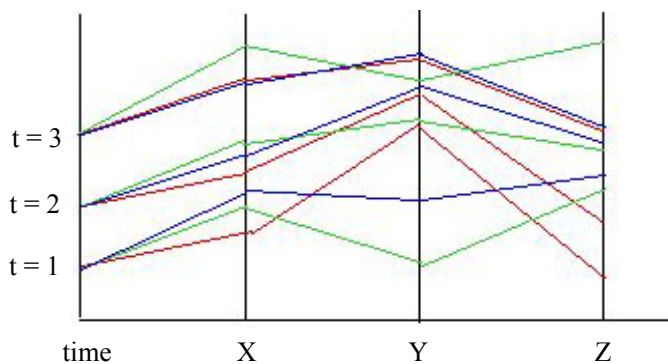


Hình 4. Takagi Sugeno Neuro-fuzzy system

2.3 Trục quan hóa

Khái niệm kết hợp song song (Parallel Coordinates) [8] được đưa ra đầu tiên bởi Maurice d'Ocagne, năm 1885, sau đó được giới thiệu rộng rãi bởi Alfred Inselberg, năm 1959 và được sử dụng như một công cụ trực quan hóa (Visualization). Thực chất đồ thị kết hợp

song song chỉ là đồ thị ở dạng thanh, trong đó có N trục Y với các miền giá trị khác nhau trải đều trên trục X, tại một thời điểm giá trị trên các trục Y được nối lại với nhau để biểu diễn mối quan hệ giữa các thành phần trên các trục Y.



Hình 5. Đồ thị song song biểu diễn mối quan hệ giữa time và X, Y, Z

3.ỨNG DỤNG KỸ THUẬT NEURO-FUZZY VÀ KỸ THUẬT TRỰC QUAN HÓA TÌM QUY LUẬT NHÂN QUẢ TRONG CÔNG THỨC DƯỢC PHẨM

3.1 Dữ liệu công thức

Dữ liệu công thức dược phẩm được tham khảo từ công trình của Gohel & Amin [3], công thức viên nén phóng thích có kiểm soát bao gồm 27 công thức thực nghiệm với 3 biến độc lập và 4 biến phụ thuộc tương ứng như sau:

Các biến độc lập	
x ₁	Tốc độ khuấy (RPM)
x ₂	Lượng CaCl ₂
x ₃	% liquid paraffin
Các biến phụ thuộc	
y ₁	Thời gian 80% hoạt chất phóng thích
y ₂	% hoạt chất phóng thích sau 60 phút
y ₃	% hoạt chất phóng thích sau 360 phút
y ₄	% hoạt chất phóng thích sau 480 phút

Một bộ dữ liệu khác được tham khảo từ công trình của Bodea và Leucuta [2], công thức viên nén matrix bao gồm 14 công thức thực nghiệm với 3 biến độc lập và 3 biến phụ thuộc như sau:

Các biến độc lập	
x ₁	% HPMC
x ₂	% CMCNa
x ₃	% HCL

Bảng 1. Dữ liệu thực nghiệm và kết quả dự đoán từ kỹ thuật neuro-fuzzy

	x ₁	x ₂	x ₃	y ₁	y _{1_pre}	y ₂	y _{2_pre}	y ₃	y _{3_pre}	y ₄	y _{4_pre}
1	500	5	25	434	469.461	42.35	39.122	73.86	70.599	80.28	79.885
2	500	5	50	453	475.936	37.26	36.824	70.84	69.581	81.77	79.13
3	500	10	0	477	475.752	40	40.765	69.29	69.628	80.31	79.103
4	500	10	25	483	489.499	37.32	38.446	69.17	68.361	77.79	78.28
5	500	10	50	499	510.37	36.82	34.625	68.36	66.816	76.26	77.167
6	500	15	0	454	494.001	40.65	42.102	70.57	67.579	81.62	77.679
7	500	15	25	532	522.706	41.84	37.258	66.84	65.255	73.54	75.407
8	500	15	50	579	579	29.71	29.71	61.82	61.82	73.07	73.07
9	1000	5	25	375	392.583	40.72	44.776	76.8	77.304	85	85.59
10	1000	5	50	385	387.386	40.84	50.194	77.43	85.108	84.86	94.764
11	1000	10	0	395	417.531	42.14	41.678	76.58	68.831	89.62	83.598

Các biến phụ thuộc	
y ₁	% hoạt chất phóng thích sau 1 giờ
y ₂	% hoạt chất phóng thích sau 6 giờ
y ₃	% hoạt chất phóng thích sau 12 giờ

Dữ liệu công thức viên nén matrix đã được tối ưu trong nghiên cứu của Bodea và Leucuta [2] bằng phương pháp thống kê D-optimal quadratic. Bài báo này sẽ trình bày kết quả các dữ liệu trên với mô hình dự đoán quan hệ nhân quả dùng kỹ thuật neuro-fuzzy và visualization, đồng thời thực hiện so sánh kết quả của máy tính với kết quả thống kê của các tác giả trên.

3.2 Công cụ phần mềm

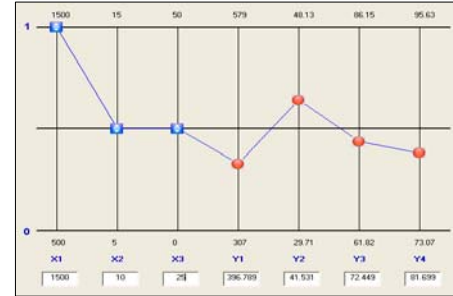
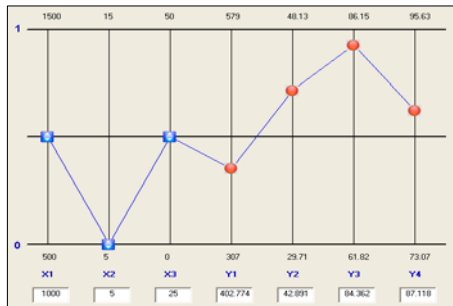
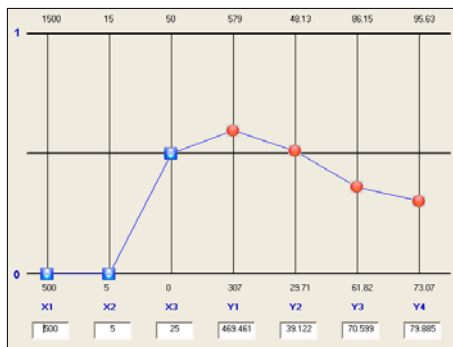
Phần mềm được thiết kế với sự kết hợp của kỹ thuật mạng neuron kết hợp với logic mờ và hiển thị kết quả qua đồ thị kết hợp song song.

4.KẾT QUẢ VÀ BÀN LUẬN

4.1. Công thức viên nén phóng thích có kiểm soát

Dữ liệu thực nghiệm từ nghiên cứu của Gohel và Amin [3] sẽ được chia thành 2 nhóm: dữ liệu luyện và dữ liệu thử, trong đó dữ liệu thử bao gồm các 3 công thức (1; 10; 19), 24 công thức còn lại được sử dụng như dữ liệu luyện.

12	1000	10	25	429	426.37	38.83	44.448	72.98	73.615	83.42	82.573
13	1000	10	50	448	438.884	37.48	49.081	71.27	81.98	82.26	80.542
14	1000	15	0	462	451.627	40.34	42.575	70.02	64.22	80.98	91.045
15	1000	15	25	475	482.486	39.98	43.722	69.41	67.672	80.49	77.423
16	1000	15	50	503	541.546	39.5	46.45	67.38	75.916	75.8	51.46
17	1500	5	25	307	359.473	48.13	40.331	86.15	65.668	93.52	77.897
18	1500	5	50	321	359.415	45.42	59.601	81.84	71.331	95.63	85.009
19	1500	10	0	348	392.895	40.32	35.902	79.25	72.082	91.75	77.863
20	1500	10	25	370	398.451	40.97	44.787	77.27	72.022	85.52	80.514
21	1500	10	50	382	405.406	41.02	59.598	77.77	72.226	85.23	84.316
22	1500	15	0	415	444.609	40.21	48.131	75.23	87.593	83.46	86.565
23	1500	15	25	436	463.216	42.25	51.977	73.69	82.607	80.09	84.791
24	1500	15	50	511	497.006	38.5	59.245	66.91	74.039	73.87	82.061



Hình 6. Đồ thị song song biểu diễn mối tương quan giữa các biến độc lập x_i và các biến phụ thuộc y_i trong công thức viên nén phóng thích có kiểm soát

Hình 6a thể hiện giá trị của các biến x_1 và y_1 theo dữ liệu dự đoán cho công thức 1, dễ dàng ta thấy được trong hình 6b, nếu ta nâng giá trị của x_1 lên mức trung bình và giữ nguyên các giá trị x_2, x_3 thì y_1 giảm, y_2, y_4 tăng và y_3 tăng rất nhanh so với dữ liệu hình 6a. Nếu ta tiếp tục tăng x_1 đến Max và vẫn giữ nguyên x_2, x_3 (hình 6c) thì so với dữ liệu hình 6b, y_1 tiếp tục giảm, y_2 giảm nhẹ, y_3 giảm nhanh và y_4 tăng nhẹ. Trong hình 6d, ta chọn x_1 là Max, x_2 và x_3 ở mức trung bình thì so với dữ liệu hình 6c, y_1, y_2 tăng nhẹ, y_3 giảm nhẹ và y_4 giảm nhiều. Vậy với công cụ hỗ trợ là đồ thị kết hợp song song người sử dụng có thể thấy rõ sự thay

đổi giá trị của các biến vào ra trong công thức khi ta thay đổi giá trị của một thành phần nào đó trong các biến này, tri thức thu được trực quan và dễ hiểu hơn so với tri thức khô khan dạng luật “If ... then ...”.

4.2 Công thức viên nén matrix

Dữ liệu thực nghiệm từ nghiên cứu của Bodea và Leucuta [2] sẽ được chia thành 2 nhóm: dữ liệu luyện và dữ liệu thử, trong đó dữ liệu thử bao gồm các 2 công thức (1; 6), 12 công thức còn lại được sử dụng như dữ liệu luyện.

Bảng 2. Giá trị R^2 của mô hình Neuro-fuzzy và phương pháp thống kê (Gohel và Amin, 1998)

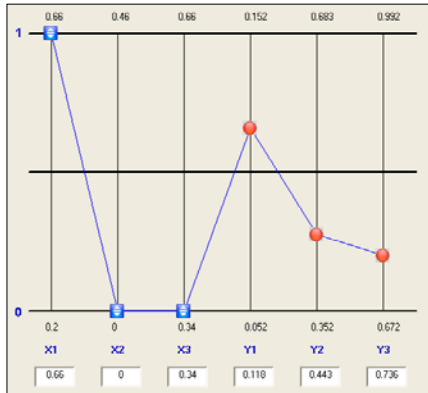
Method	t_{80}	Y_{60}	Y_{360}	Y_{480}
Neuro-fuzzy	R^2 Train= 84.98%	R^2 Train= 99.96%	R^2 Train= 99.96%	R^2 Train= 99.95%
Statistical	R^2 =99%	R^2 =74%	R^2 =95%	R^2 =92%

Bảng 3. Dữ liệu thực nghiệm và kết quả dự đoán từ kỹ thuật neuro-fuzzy

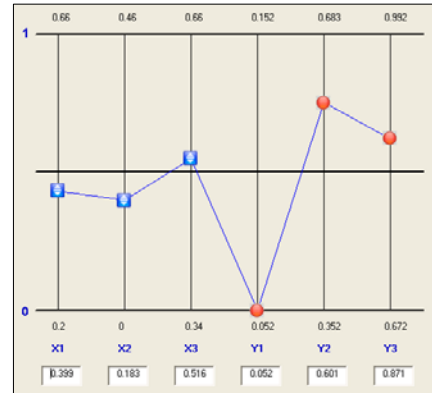
	x_1	x_2	x_3	y_1	y_{1_pre}	y_2	y_{2_pre}	y_3	y_{3_pre}
1	0.66	0	0.34	0.122	0.118	0.448	0.443	0.712	0.736
2	0.34	0	0.66	0.152	0.144	0.683	0.645	0.992	0.975
3	0.2	0.46	0.34	0.104	0.081	0.545	0.488	0.902	0.837
4	0.2	0.14	0.66	0.112	0.122	0.612	0.647	0.986	0.982
5	0.553	0	0.446	0.143	0.12	0.518	0.449	0.792	0.725
6	0.446	0	0.553	0.148	0.148	0.585	0.585	0.866	0.866
7	0.506	0.153	0.34	0.074	0.103	0.388	0.447	0.68	0.718
8	0.353	0.306	0.34	0.052	0.089	0.352	0.488	0.672	0.831
9	0.2	0.353	0.446	0.098	0.091	0.576	0.488	0.925	0.836
10	0.35	0.15	0.5	0.084	0.085	0.512	0.512	0.856	0.856
11	0.35	0.15	0.5	0.087	0.085	0.518	0.512	0.862	0.856
12	0.35	0.15	0.5	0.084	0.085	0.507	0.512	0.851	0.856

Bảng 4. Giá trị R^2 của mô hình Neuro-fuzzy và phương pháp thống kê (Bodea và Leucuta, 1997)

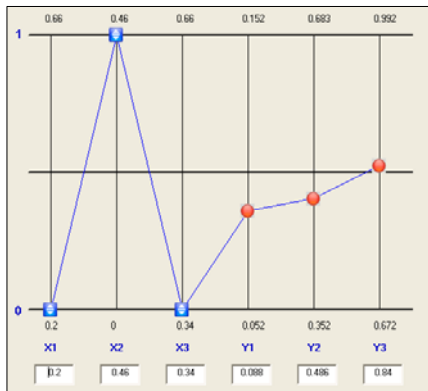
Method	Y_1	Y_2	Y_3
Neuro-fuzzy	R^2 Train= 68%	R^2 Train= 98.12%	R^2 Train= 99.33%
Statistical	R^2 =96%	R^2 =88%	R^2 =91%



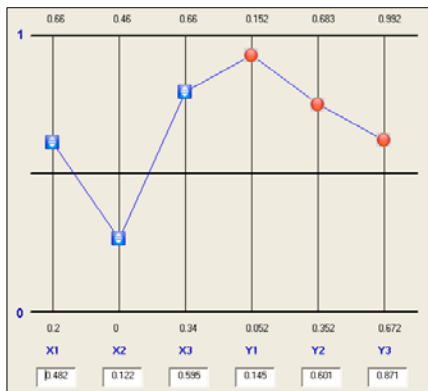
(a)



(d)



(b)



(c)

Hình 7. Đồ thị song song biểu diễn mối tương quan giữa các biến độc lập x_i và các biến phụ thuộc y_i trong công thức viên nén matrix

Tương tự cho dữ liệu công thức viên nén phóng thích có kiểm soát, đối với công thức viên nén matrix hình 7a thể hiện giá trị các biến vào ra theo công thức 1. Nếu ta giảm x_1 xuống Min, nâng x_2 lên Max, và giữ nguyên giá trị x_3 như hình 7b thì y_1 giảm nhiều, y_2 tăng ít và y_3 tăng nhiều. Tương tự như vậy khi ta thay đổi giá trị các biến đầu vào x_i như hình 7c và hình 7d thì ta dễ dàng thấy được sự thay đổi giá trị các biến đầu ra y_i và đọc được chính xác các giá trị này trên các ô giá trị tương ứng.

5. KẾT LUẬN

Mặc dù các kỹ thuật mạng neuron, logic mờ đã được giới thiệu từ lâu nhưng những ứng dụng dựa trên các kỹ thuật này vẫn được quan tâm rất nhiều. Thực nghiệm cho thấy việc ứng dụng neuro-fuzzy và visualization tìm quy luật nhân quả trong công thức được phẩm đạt được kết quả có độ chính xác cao hơn, nhanh chóng hơn so với việc thực hiện bằng phương pháp thống kê truyền thống. Ngoài ra, việc thể hiện trực quan mô hình nhân quả bằng đồ thị kết hợp song song cung cấp cho người dùng tri thức trực quan và cụ thể hơn so với tập luật "If ... then ..." và các phương pháp truyền thống trước đây.

EXTRACTING CAUSE-EFFECT RELATIONSHIPS IN PHARMACY PRODUCTS USING NEURO-FUZZY SYSTEM COMBINED TO VISUALIZATION TECHNIQUE

Nguyen Dang Khoa⁽¹⁾, Do Quang Duong⁽²⁾

(1) Dongnai Vocational College

(2) University of Medicine and Pharmacy in HCMC

ABSTRACT: *Neuro-fuzzy system is a fusion of functionalities in neural networks and fuzzy logic in order to model and extract knowledge from data. This research presents an application of neuro-fuzzy combined to visualization approach for extracting cause-effect relationships between ingredients and properties in formulation. This result will lead formulators to understanding their products more precisely and saving a lot of time and labor in R&D process.*

Keywords: *Neural networks, Neural-fuzzy, Visualization, Formulation, Controlled Release.*

TÀI LIỆU THAM KHẢO

- [1]. Đặng Văn Giáp, *Áp dụng phần mềm thông minh tối ưu hóa công thức và quy trình*, Khoa Dược – Đại học Y Dược Tp. HCM, (2001).
- [2]. Bodea, A., Leucuta, S.E., *Optimization of hydrophilic matrix tablets using a D-optimal design*, Int. J. Pharm. 153, 247-255, (1997).
- [3]. Gohel, M.C., Amin A.F., *Formulation optimization of controlled release diclofenac sodium microsphere using factorial design*, Journal of Controlled Release. 51, 115-122, (1998).
- [4]. Kiem Hoang, Bac Le Hoai, Duong Quang Do, *A Combination Approach of AI Technics for Optimization Problem in Pharmaceutics*, the 7th World Multiconference on Systemics, Cybernetics and Informatics, (2003)
- [5]. J.R. Jang, C. Sun and E. Mizutani, *Neuro-fuzzy and Soft Computing, A Computation Approach To Learning And Machine Intelligence*, Prentice Hall, Upper Saddle River, (1997).
- [6]. Lin, C.T and Lee, C.S.G, *Neural Fuzzy Systems, A Neuro-Fuzzy Synergism to Intelligent Systems*, Prentice Hall International, (1996).
- [7]. P.P. Bonissone, *Adaptive Neural Fuzzy Inference Systems (ANFIS): Analysis and Applications*, GE CRD, Schenectady, NY USA, (1997).
- [8]. Computational Engineering Systems Lab, *Center for Computer Aids to Industrial Productivity*, Directed by Richard L. Peskin, <http://www.caip.rutgers.edu/~peskin/epriRpt/index.html>.