

**ỨNG DỤNG MẠNG NƠN TỰ TỔ CHỨC  
 ĐỂ ĐÁNH GIÁ THỊ TRƯỜNG ĐÓNG TÀU THẾ GIỚI**  
 USING SELF-ORGANIZING MAP (SOM) NEURAL NETWORK TO EVALUATE  
 WORLD SHIPBUILDING MARKET

**TS. LÊ VĂN ĐIỂM**  
*Khoa Máy tàu biển, Trường ĐHHH*

**Tóm tắt**

*Bài báo giới thiệu việc ứng dụng mạng nơon nhân tạo tự tổ chức (SOM) để phân tích dữ liệu thống kê đa chiều. Dữ liệu thống kê về thị trường đóng tàu thế giới được sử dụng làm ví dụ minh họa cho việc áp dụng thuật toán SOM để phân tích dữ liệu.*

**Abstract**

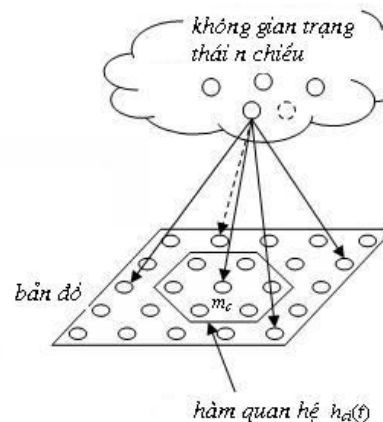
*In the paper, the model of SOM neural network is introduced to analyze multi-variable data. A set of world shipbuilding data is used as an example of SOM to visually analyze multi-variable data.*

**1. Đặt vấn đề**

Việc giải quyết nhiều bài toán thực tế thường liên quan đến xử lý các mảng dữ liệu đa chiều. Việc sử dụng các công cụ toán học truyền thống để phân tích, xử lý dữ liệu đa chiều lại gặp nhiều khó khăn, đặc biệt khi dữ liệu có nhiều biến và số mẫu dữ liệu lớn. Mạng nơon nhân tạo là một công cụ tính toán hiện đại đang được phát triển, mà việc áp dụng nó tỏ ra có nhiều lợi thế so với các phương tiện tính toán truyền thống. Ngày nay, cùng với sự phát triển mạnh mẽ của các thế hệ máy tính mới, công nghệ mạng nơon nhân tạo ngày càng được ứng dụng rộng rãi và mang lại hiệu quả cao trong nhiều lĩnh vực khoa học.

**2. Thuật toán SOM phân tích dữ liệu đa chiều**

*Bản đồ tự tổ chức* (Self-organizing Maps - SOM) là một dạng mạng nơon sử dụng thuật toán huấn luyện không điều khiển (unsupervised training) gồm có một lớp vào và một lớp ra. Thuật toán SOM được dùng để ánh xạ không gian dữ liệu đa chiều xuống môi trường một, hai hoặc ba chiều (thường áp dụng hai chiều). Nhờ vậy việc đánh giá và xử lý mối quan hệ giữa các biến trong không gian hai chiều trở lên dễ dàng hơn. Giả sử không gian dữ liệu được đặc trưng bởi tổ hợp  $n$  biến  $x = [x_1, x_2, \dots, x_n]$ , nghĩa là vectơ trạng thái trong không gian  $n$  chiều. Khi áp dụng SOM các vectơ này sẽ tự sắp xếp lên mặt phẳng theo một tiêu chuẩn tương tự nào đó, thường được sử dụng là khoảng cách Euclidean. Có nghĩa là mỗi vectơ trạng thái sẽ tương ứng với một điểm tọa độ  $m_i = [m_{i1}, m_{i2}, \dots, m_{in}]$  trên bản đồ trạng thái (Hình 1),  $m_i$  được gọi là *véctơ mô hình* hay *nơon*.



**Hình 1. Thuật toán SOM nhận dạng TTKT.**

Việc huấn luyện SOM được thực hiện bằng cách truyền các vectơ trạng thái tới lớp vào của mạng. Thuật toán SOM sẽ hiệu chỉnh vị trí của các nơon sao cho khoảng cách từ chúng đến các vectơ trạng thái là nhỏ nhất. SOM sử dụng thuật toán huấn luyện lặp như sau [3]:

1. Hình thành ngẫu nhiên các vectơ mô hình  $m_i$  (tọa độ của các nơon trên bản đồ);
2. Xác định nơon  $m_c$  có khoảng cách nhỏ nhất đến vectơ trạng thái (gọi là *nơon chiến thắng - neuron-winner*), mà chỉ số của nó thỏa mãn điều kiện:

$$\|x(t) - m_c(t)\| = \min \{\|x(t) - m_i(t)\|\}, \quad (1)$$

thường áp dụng khoảng cách Euclidean  $\|x - m\| = \sqrt{\sum_{k=1}^n (x - m_k)^2}$  ;

3. Hiệu chỉnh vị trí của nơron chiến thắng và các nơron lân cận trong giới hạn của hàm quan hệ  $h_{ci}(t)$  sao cho chúng càng gần với vectơ trạng thái hơn:

$$m_i(t+1) = m_i(t) + \alpha(t)h_{ci}(t)[x(t) - m_i(t)], \quad (2)$$

trong đó  $\alpha(t)$  – tốc độ huấn luyện ( $0 < \alpha(t) < 1$ ), giảm dần theo sự tăng của số lần lặp  $t$ ;  $h_{ci}(t)$  – hàm quan hệ, cũng giảm dần khi  $t$  tăng;

4. Kiểm tra điều kiện dừng, nếu chưa thỏa mãn thì quay về bước 2.

Điều kiện dừng được coi là thỏa mãn nếu lượng hiệu chỉnh (thành phần thứ hai trong công thức (2)) cho các nơron nhỏ đến mức quy định. Khi này các vectơ trạng thái sẽ rơi vào cùng một vị trí trên bản đồ khi chuyển từ vòng lặp này sang vòng lặp tiếp theo.

Kết thúc quá trình huấn luyện, bản đồ đã huấn luyện là hình ảnh của các vectơ dữ liệu  $n$  chiều trên mặt phẳng. Tiếp theo, các nơron có thể được nhóm lại thành từng nhóm, sử dụng một trong số các phương pháp phân nhóm. Như vậy theo đặc tính giống nhau của các vectơ dữ liệu, chúng được chia ra thành các phân vùng trên bản đồ. Nhờ đó có thể dễ dàng phân tích, đánh giá thuộc tính của các nhóm dữ liệu này.

### 3. Phân tích dữ liệu thị trường đóng tàu thế giới sử dụng SOM

Bảng dưới mô tả dữ liệu thống kê thị trường đóng tàu đối với một số nước chủ yếu theo số lượng tàu được đóng, chủng loại tàu và số tấn trọng tải [4].

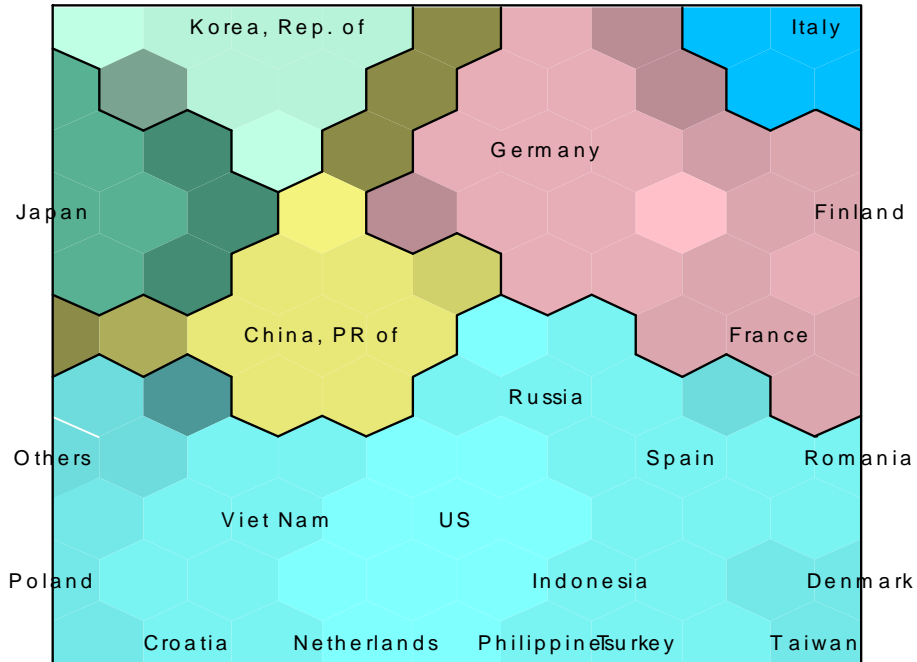
**Bảng 1. Thống kê dữ liệu thị trường đóng tàu thế giới năm 2005.**

Quốc gia	Tàu dầu		Tàu hàng rời		Tàu container		Tàu hàng tổng hợp		Tàu khách		Tổng	
	Số lượng	Số tấn	Số lượng	Số tấn	Số lượng	Số tấn	Số lượng	Số tấn	Số lượng	Số tấn	Số lượng	Số tấn
Korea	840	26289	30	614	420	16308	48	1129	1	1	1339	44341
Japan	416	11154	481	9459	93	3534	215	4369	4	28	1209	28544
China	451	9039	217	3916	336	5666	173	1576	3	61	1180	20259
Germany	16	249	4	29	145	2168	31	445	13	964	209	3855
Italy	24	188	0	0	0	0	6	64	44	2669	74	2922
Poland	4	117	0	0	39	912	46	819	4	30	93	1879
Taiwan	0	0	4	122	51	1492	1	8	0	0	56	1621
Turkey	110	870	7	46	40	439	34	182	0	0	191	1536
Croatia	39	818	0	0	0	0	28	620	3	14	70	1453
Finland	0	0	0	0	0	0	1	20	15	1345	16	1365
Romania	33	359	1	23	22	727	20	84	5	41	81	1236
Viet Nam	11	158	28	406	10	89	53	495	0	0	102	1147
Philippines	0	0	33	535	12	311	6	157	2	6	53	1008
Denmark	0	0	0	0	14	943	0	0	0	0	14	943
France	4	195	0	0	0	0	0	0	6	674	10	869
Russia	34	515	0	0	0	0	31	209	4	93	69	818
Netherlands	12	59	0	0	18	161	88	476	7	26	125	722
USA	12	296	0	0	0	0	10	357	7	68	29	721
Spain	12	192	0	0	6	77	22	141	7	176	47	586

Indonesia	8	59	18	255	0	0	32	159	7	52	65	525
Others	35	421	34	521	16	256	145	943	43	207	273	2347

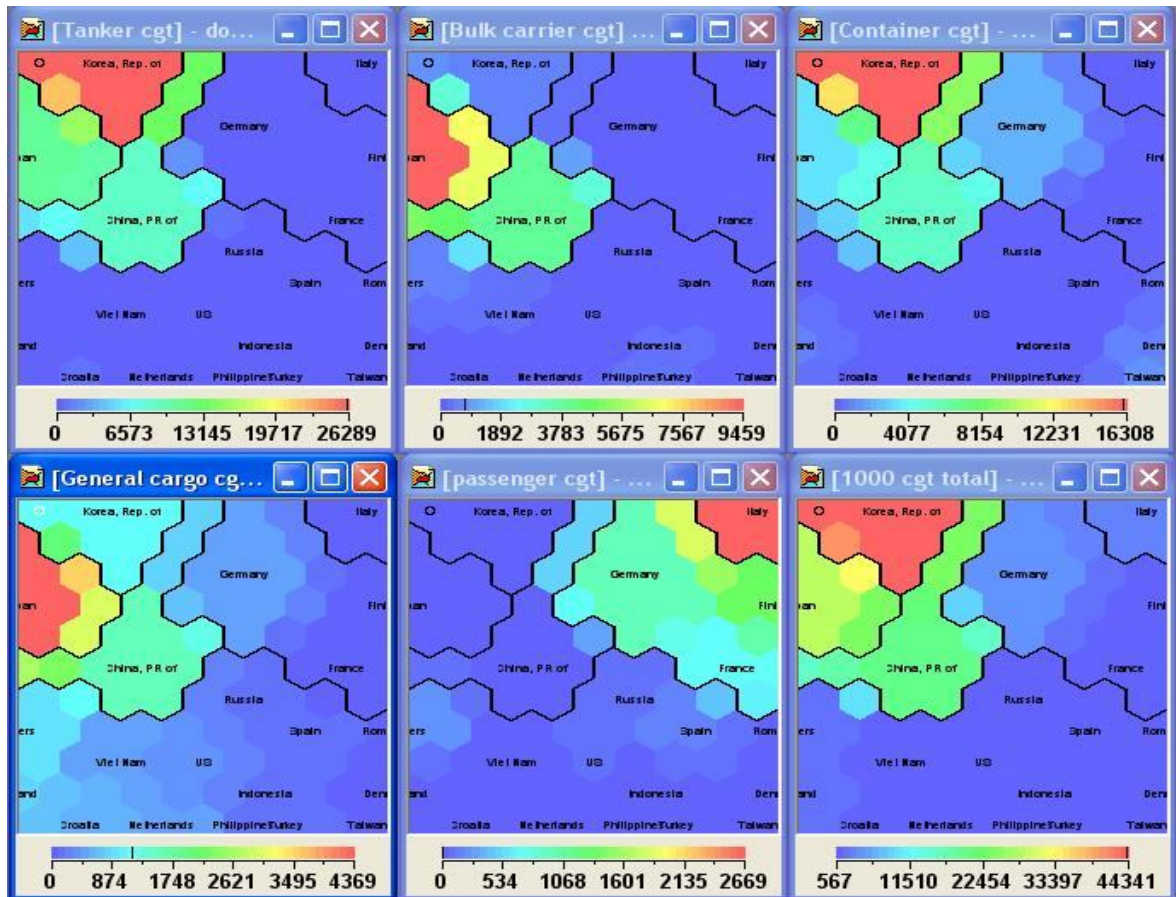
Kết quả sử dụng SOM để phân tích bảng dữ liệu trên (Hình 2) chỉ ra rằng, thị trường đóng tàu thế giới được chia thành sáu nhóm. Trong đó, ba quốc gia đang dẫn đầu thị trường đóng tàu thế giới là Hàn Quốc, Nhật Bản và Trung Quốc được nhóm thành ba nhóm gần nhau.

Clusters - dong\_tau\_minh.som



Hình 2. Bản đồ SOM thị trường đóng tàu thế giới năm 2005.

Thuật toán SOM còn cho phép xây dựng các bản đồ theo từng biến dữ liệu (components), trong ví dụ trên là chủng loại tàu. Bản đồ theo chủng loại tàu (Hình 3) cho phép dễ dàng đánh giá quan hệ giữa các biến dữ liệu (chủng loại tàu). Phần phía dưới các bản đồ theo chủng loại tàu có các thang chia mà số lượng dữ liệu thay đổi theo màu (vùng màu đỏ tương ứng với giá trị biến lớn nhất). Nhờ đó giúp dễ dàng so sánh, đánh giá giữa các biến dữ liệu.



Hình 3. Bản đồ các thành phần thị trường đóng tàu thế giới.

Từ Hình 3 cho thấy, Hàn Quốc đang dẫn đầu thị trường đóng tàu, đặc biệt là tàu dầu (tankers) và tàu chở container, trong khi Nhật Bản lại chiếm vị trí dẫn đầu ở thị trường tàu hàng rời trọng tải lớn (bulk carriers), còn Ý thì chú trọng đến đóng tàu khách.

#### 4. Kết luận

Việc ứng dụng mạng nơron nhân tạo để phân tích và xử lý dữ liệu đã cho thấy nhiều ưu điểm nổi bật. Ưu điểm của thuật toán SOM là không cần biết trước số nhóm (clusters), mà dựa trên việc phân tích các đặc điểm của tập dữ liệu, thuật toán sẽ tự phân nhóm chúng (unsupervised training). Việc sử dụng các bản đồ thành phần cho phép so sánh một cách trực quan theo từng biến của tập dữ liệu đa chiều.

#### TÀI LIỆU THAM KHẢO

- [1] Lê Văn Điềm. *Các mô hình và thuật toán chẩn đoán kỹ thuật động cơ diesel tàu thủy trong điều kiện khai thác*. Luận văn tiến sỹ kỹ thuật, Trường Đại học Giao thông đường thủy Xanh-Petecbua, Liên bang Nga, 2006 - 153 trang.
- [2] Principe J.C., Euliano N.R., Lefebvre W.C. *Neural and adaptive systems: Fundamentals through simulations*. John Wiley & Sons, 2000. – 656 p.
- [3] Treuvo Kohonen. *The Self-Organizing Maps*. Springer, 2001. - 501 p.
- [4] *Shipping statistics year book 2005*. Institute of Shipping Economics and Logistics, 2005. 470p.

Người phản biện: PGS. TS. Lê Văn Học